

REFERENCE 1 (bases 1 to 1356)
 AUTHORS Cao, J., Shapleigh, J., Gennis, R., Revzin, A. and Ferguson-Miller, S.
 TITLE The gene encoding cytochrome c oxidase subunit II from Rhodospirillum rubrum: comparison of the deduced amino acid sequence with sequences of corresponding peptides from other species
 JOURNAL Gene 101 (1), 133-137 (1991)
 MEDLINE 91285423
 PUBMED 1648008
 REFERENCE 2 (bases 1 to 1356)
 AUTHORS Hiser, C. and Ferguson-Miller, S.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUL-2001) Biochemistry Department, Michigan State University, East Lansing, MI 48824, USA
 REMARK Sequence update by submitter
 COMMENT On Jul 26, 2001 this sequence version replaced gi:15022158.
 FEATURES
 source location/Qualifiers
 1. 1356
 /organism="Rhodospirillum rubrum"
 /db_xref="taxon:1063"
 gene 349..1260
 /gene="coxII"
 CDS 349..1260
 /gene="coxII"
 /EC_number="1.9.3.1"
 /codon_start=1
 /transl_table=11
 /product="cytochrome c oxidase subunit II"
 /protein_id="AA26100.3"
 /db_xref="GI:15022157"

BASE COUNT 239 a 447 c 437 g 233 t
 ORIGIN

alignment_scores:
 Quality: 165.00 Length: 42
 Ratio: 4.125 Gaps: 0
 Percent Similarity: 95.238 Percent Identity: 73.810

alignment_block:
 US-09-712-768-4 x M57680 ..
 Align seg 1/1 to: M57680 from: 1 to: 1356

1 Proleugluilevaltrphtlilevalprovalvalillelevalphe11 17
 |||||
 643 CCGCTCGAGTGCCTGCGAGATCGTCGATCCTCGTGGCAT 692
 17 eglialapheserleuprovalleuphelsglnnglnupheproglug 34
 |||||
 693 CGGGCCCTTCTCGCTCGCTGCTTCAACGACGAGAAATCCCGGAG 742
 34 lvsplilevalilleasnaValglugly 42
 ::|||
 743 CGGACGTGACGAGTGTACGAGGC 768

seq_name: gb_ba:PDOX12
 seq_documentation_block:
 LOCUS PDOX12 1653 bp DNA BCT 21-MAR-1995
 DEFINITION P. denitrificans oxi12 gene for cytochrome c oxidase subunit II (EC 1.9.3.1).
 ACCESSION X05934
 VERSION X05934.1 GI:45483
 KEYWORDS cytochrome c oxidase subunit II.
 SOURCE Paracoccus denitrificans.
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum rubrum group;

REFERENCE 1 (bases 1 to 1653)
 AUTHORS Ludwig, B.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-1987) Ludwig B., Institut fuer Biochemie, Medizinische Universitaet zu Luebeck, Ratzeburger Allee 160, 2400 Luebeck 1
 REFERENCE 2 (bases 1 to 1653)
 AUTHORS Stehrucke, P., Steffens, G.C., Pankus, G., Buse, G. and Ludwig, B.
 TITLE Subunit II of cytochrome c oxidase from Paracoccus denitrificans.
 JOURNAL DNA sequence, gene expression and the protein
 MEDLINE Eur. J. Biochem. 167 (3), 431-439 (1987)
 PUBMED 88004464
 FEATURES
 source location/Qualifiers
 1. 1653
 /organism="Paracoccus denitrificans"
 /strain="wild type"
 /db_xref="taxon:266"
 sig_peptide /clone_1lb="pUC8"
 CDS 532..615
 /clone="470"
 /product="signal peptide (AA -28 to -1)"
 532..1425
 /note="precursor polypeptide (AA -28 to 269)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA29372.1"
 /db_xref="GI:45484"
 /db_xref="SPTREMBL:059663"
 /translation="MAITKRGVAAVNSLQVATWTPVAPALQVLDLPVIGKPVNG
 GNFPDASSPLAHDOQMDHRYVLTITAVTTFVCLLICTVRNRANVPARFTIN
 TLEIVATVLPVLLIIVAGFSLPLTFERSQMPDPDLVIKIGHQWMSVEYNDNF
 AFDALMLEKALADAGYSEDEVLATIDMPVVVPVGRKVLVQVATDVTHMTIPAFV
 KDDAVGRTAQLMFESVDOGVYFGQSELGINHAYMPIYKVAQSEKYEAMLAGAKE
 EFADASDYLPAPSPVKLASAE"
 616..1422
 /product="mature cytochrome oxidase subunit II (AA 1-269)"

BASE COUNT 270 a 525 c 556 g 302 t
 ORIGIN

alignment_scores:
 Quality: 139.50 Length: 43
 Ratio: 3.577 Gaps: 1
 Percent Similarity: 90.698 Percent Identity: 60.465

alignment_block:
 US-09-712-768-4 x PDOX12 ..
 Align seg 1/1 to: PDOX12 from: 1 to: 1653

1 Proleugluilevaltrphtlilevalprovalvalillelevalphe11 17
 |||||
 841 CCGATGAGGTATCTGACCTGCTCGGATCGATCGTGGCAT 890
 17 eglialapheserleuprovalleuphelsglnnglnupheproglug 34
 |||||
 891 CGGGCCCTTCTCGCTCGCTGCTTCAACGACGAGAAATCCCGGAG 940
 33 lvsplilevalilleasnaValglugly 42
 ::|||
 941 ATCCGACCTGTGATCAAGGCTATCGGC 969

seq_name: gb_ba:PDOX01
 seq_documentation_block:
 LOCUS PDOX01 4890 bp DNA BCT 12-SEP-1993
 DEFINITION Paracoccus denitrificans cytochrome c oxidase CoII and CoII genes
 ACCESSION X05828
 VERSION X05828.1 GI:45468
 KEYWORDS CoII gene; cytochrome aa3; cytochrome c oxidase; unidentified
 reading frame.

SOURCE Paracoccus denitrificans.
ORGANISM Paracoccus denitrificans.
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Paracoccus.
REFERENCE 1 (bases 1 to 4887)
AUTHORS Raitio, M., Jalli, T. and Saraste, M.
TITLE Isolation and of the genes for cytochrome c oxidase in Paracoccus denitrificans
JOURNAL EMBJ 3, 6, 2825-2833 (1987)
REFERENCE 2 (bases 1 to 4890)
AUTHORS Saraste, M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1988) to the EMBL/Genbank/DBJ databases
COMMENT Data kindly reviewed (07-MAR-1988) by Saraste M.
FEATURES
source
1..4890
Location/Qualifiers
/organism="Paracoccus denitrificans"
/db_xref="taxon:266"
/map="cox1 locus"
853..856
old_sequence
/note="gcc was gcc in [1]"
/citation=[1]
862..863
old_sequence
/note="ag was a in [1]"
/citation=[1]
880..882
old_sequence
/note="agt was at in [1]"
/citation=[1]
892..894
old_sequence
/note="tgg was tg in [1]"
/citation=[1]
898..900
old_sequence
/note="gcc was gc in [1]"
/citation=[1]
900..906
repeat_unit
/note="inverted repeat A"
912..915
old_sequence
/note="TATA-box like sequence"
924
old_sequence
/note="g was c in [1]"
/citation=[1]
937..938
old_sequence
/note="tg was tccg in [1]"
/citation=[1]
940..946
repeat_unit
/note="inverted repeat A'"
984..987
RBS
/note="pot. ribosome binding site"
992..1888
CDS
/note="coli gene product (AA 1 - 298)"
/codon_start=1
/transl_table=11
/protein_id="CAA29268.1"
/db_xref="GI:45469"
/db_xref="SWISS-PROT:P08306"
/translation="MMAIATKRGVAAVNSLGVATWTAVPALADVLGDLPIVIGKPVN
GNNFQPSAPLADDOOMDHEFLYITATVTFVCLLLICIVFRNRPANVPARTH
NPPIEVITWLVPLVILVIGAFSLPIIFPSOEMPNDPLVIAIGAGVWVSEYEPDG
VARDALMLEREALADAGSDEDEYLLATDNDPVVPGKKVLYVOTATDVIHAWTIPAF
VKODAVPGRTAQLMESVDDEGYVFGCCSELGINAHYIMFIVYKASQEKTEAMLAGAK
EEFPAADSGRLPSPVKLASAE"
1961..1965
RBS
/note="pot. ribosome binding site"
1974..2813
CDS
/note="orf 1 (AA 1 - 279)"
/codon_start=1
/transl_table=11
/protein_id="CAA29269.1"
/db_xref="GI:45470"
/db_xref="SWISS-PROT:P08301"
/translation="MSLVVTFATVGLMIAQPPVNPFAVCALFLIAGGASGALNMW
YDADIDAVMRTAGRPVPSGRVTSQRIALGIALGSLVNMIGAGGNMFRAGFLAFTI
FFAVVITTWLKRSTPONTIVIGAGAGFPFMIGMALPTGGIGIESLIMRALIFPWPFP

HFWALALFMKKDYSKAGVEMLVTVNKRKVTCHRIEAYTLVLAPALMLGFTSVGPLY
LAVSVLNLALFIAGWQILRRSEDOAQADGYVEVKRYFRLSLYTFHLHALLVQHWV
GSW"
2802..2806
RBS
/note="pot. ribosome binding site"
2813..3082
CDS
/note="orf 2 (AA 1 - 89)"
/codon_start=1
/transl_table=11
/protein_id="CAA29270.1"
/db_xref="GI:45471"
/db_xref="SWISS-PROT:P08302"
/translation="MLQVEHLKRRSRNIGLLVLLAFALVFGLSVVKITQGDGM
MGYDHRPRASMLPPDDPPAPNAAVPAAGTPAAQGVTOEANO"
3079..3663
CDS
/note="orf 3 (AA 1 - 194)"
/codon_start=1
/transl_table=11
/protein_id="CAA29271.1"
/db_xref="GI:45472"
/db_xref="SWISS-PROT:P08303"
/translation="MSGKPRSNTRVYMLAGVVLMLGALSMNAAPFYSFWCKVTGFA
CTTNVAAASDVYIDKIRVRRPANDSNIGTFRRMORREMLKIGENALIAFEALNN
TDEPVGTASTYNVAPDAGYFPNKKIECFCTBOTLPDGEVEKVPVFPVADADLVNDRD
AVIRDTLSYTFHRTDPPAPKQALDKTEPTVN"
3686..3691
RBS
/note="pot. ribosome binding site"
3698..4522
CDS
/note="coli gene product (AA 1 - 274)"
/codon_start=1
/transl_table=11
/protein_id="CAA29272.1"
/db_xref="GI:45473"
/db_xref="SWISS-PROT:P06030"
/translation="MAHVKNNDYQIILPPSIIMPERGATGAVMLTGAVAMKGLTPFGL
EVEGPMELIGVLVLYVNRGMADYVNEGTGEHNPVRIQXGFIILFIMSEVMF
VAMPWAFIKNALYPMGPDSPDKGWPPEBGLVTFDPNHLPLINTLILLSGVAYTNAN
HAFVLEGTDRKTTINGLIYAVILVGCFTGLDAYEYSHAPELADLVAGATYMGREG
AHVILGTFILFVCLIRLKGQMTOKOHVGEAAMVWHFVDVWVLELVVITYMRG"
4543..4549
repeat_unit
/note="inverted repeat B"
4543..4574
terminator
/note="pol. transcription terminator"
4559..4565
repeat_unit
/note="inverted repeat B'"
BASE COUNT 777 a 1581 c 1623 g 909 t
ORIGIN

alignment_scores: Length: 43
 Quality: 139.50 Gaps: 1
 Ratio: 3.577
Percent Similarity: 90.698 Percent Identity: 60.465

alignment_block: ..

Align seg 1/1 to: PDCOX1 from: 1 to: 4890

```

1  Prolengluilevaltrprhrllevalprovalvalilleuvalphei 17
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1304 CCGATCGAGGTGATCTGGACCTGGTCCGGTGTATCTGTGGCGAT 1353
   17  eglYAlapheserleuProvalleuphetysgInglIngluPhePro...G 33
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1354 CGGGGCGCTCTCGCTCGCGATCGCTGCCGACGACGAGATGATCCGACG 1403
   33  luglyAspIlevalIleasnValGluclly 42
   ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
1404 ATCCGACCTGGTGTATCAAGCGCATTCGGC 1432
   seq_name: gb_ba:AE006001
```

```

seq_documentation_block:
LOCUS      AE006001      1188 bp      DNA      BCT      28-MAR-2001
DEFINITION Caulobacter crescentus section 327 of 359 of the complete genome.
ACCESSION AE006001 AE005673
VERSION    AE006001.1 GI:13425117
KEYWORDS
SOURCE
ORGANISM   Caulobacter crescentus.
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE  1 (bases 1 to 1188)
            Nieman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,
            Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R.,
            Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D.,
            Ely, B., Debroy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
            Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J.,
            Berry, K., Uitterback, T., Tran, K., Wolf, A., Vamathevan, J.,
            Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and
            Fraser, C.M.
            Complete genome sequence of Caulobacter crescentus
            Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
TITLE      2 (bases 1 to 1188)
JOURNAL    Nieman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,
MEDLINE    Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I.,
REFERENCE  Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
AUTHORS     Laub, M.T., Debroy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
            Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouli, H., Shetty, J.,
            Berry, K., Uitterback, T., Tran, K., Wolf, A., Vamathevan, J.,
            Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
            Fraser, C.M.
            Direct Submission
            Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source     1..1188
            /organism="Caulobacter crescentus"
            /db_xref="taxon:69394"
            /complement(94..1752)
            /gene="CC3406"
            /complement(94..1752)
            /gene="CC3406"
            /note="identified by match to protein family HMM"
            /codon_start=1
            /transl_table=1
            /product="cytochrome c oxidase, subunit I"
            /protein_id="AAK25368.1"
            /db_xref="GI:13425118"
            /translation="MASADYTHDAHDHDKPPFANFSTNHKIDITLYILFALMA
            GIVGASGLIRWEIMAPGIVQVSDTGMLOMGIFFKRGHYNVNTAHALIMTFEVM
            PAMIGFGNMFVPIMGAPDMAPERNNNISFWLLIAMILLVMEFVADGPGHFGG
            WTIYPLSTIGHKGPAMDIALSLHLGASSILGAINETITFMRAFGWTLHMPLE
            VWSYLTAFLLLSLPYLAGAITMLLTRNMGTHFPDPAGDPPVMEOHLEFMEFGE
            VYILIPGFGIISHIVSTFSKRPVEGFLAMAVAMVLAIGVFVNAHHMYTGMSTNL
            RAYVAATMIIVPTGVKIFPSMATMGSGSTSKPTMLNIGITPIFTVGAGVGVLS
            NAGIDYLTHTYVVAHFHYVLSGAVFAITFAGTITWEEKMVCVKNELGAAHEWIM
            PIGVNLITFFQHFLGQGMRRYVDPAFTLNNVSVYGVMTITVVGVGVMFLITEA
            AIRRKAENPWEAGATLEWLTSPPPHFQSEPPVIXKDDHH"
            /complement(1779..2708)
            /gene="CC3407"
            /complement(1779..2708)
            /gene="CC3407"
            /note="identified by match to protein family HMM"
            /codon_start=1
            /transl_table=1
            /product="cytochrome c oxidase, subunit II"
            /protein_id="AAK25369.1"
            /db_xref="GI:13425119"
            /translation="MKAFSGSIRIRITGASALATMMPAGHALAEDLMQOPTGADL
            QPASELKHAIWFHMLIMPITATLTLVLGLLIIVRYNKNKAEDVPAKSHNTTI
            EYIMTVVAVLILVAIFSRILFHNPKRPDLTKATGYOMGYEYDPOKRSII
            SLPTKEQAEAKKVPYLLAADNANVVPKRVAVOVTAADVIAHAFILPFGIKTDAIP
            GRINTEMKADRTGVYGGCSELGVDHAFMPEIETVYVQAEFDANVAKKTAPPAPAA

```

```

gene       3083..3409
            /gene="CC3408"
            /complement(3409)
            /gene="CC3408"
            /note="identified by Glimmer2; putative"
            /codon_start=1
            /transl_table=1
            /product="conserved hypothetical protein"
            /protein_id="AAK25370.1"
            /db_xref="GI:13425120"
            /translation="MPETIEIOLKGVIALCVALLSHADSYAYETIASLTIDIDMGE
            GTIYPLMRNMQSGIVETIYVESQSGPPKRYITUTRAGRESLAQKAEMTATRTAVNN
            IVAA"
            /complement(3699)
            /gene="CC3409"
            /complement(3699)
            /gene="CC3409"
            /note="identified by Glimmer2; putative"
            /codon_start=1
            /transl_table=1
            /product="hypothetical protein"
            /protein_id="AAK25371.1"
            /db_xref="GI:13425121"
            /translation="MTRAFETRLRGLSGLPATRTADIADHEAFSDQAAGTTEA
            EVAAIGDPPRLARETRARAGLKKOMEKTPSNAAATITALLGLAIDIVITPLILMG
            VTAGTGLVAVYITTFVGGFIFAGGPIAAGPGLTAPLTGILMASATFTGSLTLT
            VTAGTGLVAVYITTFVGGFIFAGGPIAAGPGLTAPLTGILMASATFTGSLTLT
            LEQA"
            /complement(4802)
            /gene="CC3410"
            /complement(4802)
            /gene="CC3410"
            /note="identified by Glimmer2; putative"
            /codon_start=1
            /transl_table=1
            /product="hypothetical protein"
            /protein_id="AAK25372.1"
            /db_xref="GI:13425122"
            /translation="MSRECTMLIRNTLIVAGSALVAVCGRAGVAAIAGELMTRK
            TYPGDSASVYTTDDNGKIVRYVHAGAGAPSPVWDTLTPWTKDITLIDIPVYV
            VOGAEKATIVSGPKSTVDRKAVEGRISLDDINLADGAILMGLPGDLQVRSRDMKI
            TVADPVTRFEVRSGLDIDIRYEQALANLIDVSGDVEAEKGVETTVENSQGLAD
            LQDLKRAEALIEVSGSGAVPATKAKIDISGSDVELRTPKSVTSETIGSGEVR
            EY"
            /complement(4964)
            /gene="CC3411"
            /complement(4964)
            /gene="CC3411"
            /note="identified by match to PFAM protein family HMM
            /codon_start=1
            /transl_table=1
            /product="ltd protein"
            /protein_id="AAK25373.1"
            /db_xref="GI:13425123"
            /translation="MTSAFGSLTCTNNAPPIPASSLTLDAAGVDDPBARALIGEALA
            GADGDELFAFERESAPFVDDGRKLSAYSDGEGGLRVAGCEPAGYAHAEISEPAI
            RASPSALISAKRGKGVSAEGPRATNEKLYEDNVPASPDSDKVALLOEIDARAR
            DRRVYVYVNASMAGERVVEILRADGRLLIRVPLVIRVAVQVYVERKDRREGSSGAG
            RAGFAWISPDQDQVDEALMALVNLDAVACPAGEDVVLGGMNGVLLHAEVAGH
            LEGDPNRKGISAFSGRIGERYVADPVYFDGDSILPGRGSLTVDEGTPRTTLIED
            GLVGYMDRMSARLMGKATNGRROSAYAHMPMRMTNTGMLAGEHDPOMIASLR
            GLVGYMDRMSARLMGKATNGRROSAYAHMPMRMTNTGMLAGEHDPOMIASLR
            NDPDFDPGIGVCGKSGGVGVGQPSLKITGLTVGGTAV"
            /complement(6601)
            /gene="CC3412"
            /complement(6601)
            /gene="CC3412"
            /note="identified by match to protein family HMM"
            /codon_start=1
            /transl_table=1
            /product="acetyltransferase, GNAI family"
            /protein_id="AAK25374.1"

```


CDS	/gene="coxC" 3314..4303 /gene="coxC" /codon_start=1 /transl_table=11 /product="heme O synthase" /protein_id="CA61745.1" /db_xref="GI:1177575" /db_xref="SPTREMBL:O51322"		
BASE COUNT	749 a	1252 c	1415 g 964 t
alignment_scores:	Quality: 99.00 Ratio: 2.676 Percent Similarity: 88.095	Length: 42 Gaps: 0 Percent Identity: 40.476	
alignment_block:	US-09-712-768-4 x NMCOXABC ..		
Align seg 1/1 to: NMCOXABC from: 1 to: 4380			
2 LeuGluIleValIThrIleValProValIValIleLeuValPheIleGl 18	1039 ATCGAGGTTGCGTGCATCCGTTTCGATCCGTCGCGCATTCG 1088		
18 ValAlaPheSerLeuProValIleuPheLysGlnGlnIleuPheProGluGly 35	1089 GGTTCCCTCATTCGCTTTTCTTTTGACACGACTTGACCTTCGAAGCGTG 1138		
35 spLIleValIleAsnValGluGlyArg 43	1139 ACCTGACGCTGAAGCGACCGCGCAAG 1164		
seq_name: gb_ba:RPX02			
seq_documentation_block:			
LOCUS RPX02 312430 bp DNA BCT 12-NOV-1998			
DEFINITION Rickettsia prowazekii strain Madrid E, complete genome; segment 2/4.			
ACCESSION AJ235271 AJ235269			
VERSION AJ235271.1 GI:3868717			
KEYWORDS complete genome.			
SOURCE Rickettsia prowazekii.			
ORGANISM Rickettsia prowazekii.			
REFERENCE 1 (bases 1 to 312430)			
AUTHORS Andersson, S.G.			
TITLE The genome sequence of Rickettsia prowazekii and the origin of mitochondria			
JOURNAL Nature 396 (6707), 133-140 (1998)			
MEDLINE 99039499			
REFERENCE 2 (bases 1 to 312430)			
AUTHORS Andersson, S.G.E.			
TITLE Direct Submission			
JOURNAL Submitted (11-NOV-1998) S.G.E. Andersson, Siv. Andersson@molbio.uu.se, Dept. of Molecular Biology, University of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN			
COMMENT On Nov 13, 1998 this sequence version replaced g1:3860788.			
FEATURES			
Source	1..312430		
	/organism="Rickettsia prowazekii"		
	/strain="Madrid E"		

TITLE

Direct Submission
Submitted (20-MAY-1999) Rossmann R., Eidgenossische Technische
Hochschule Zuerich, Mikrobiologisches Institut, Schmelzbergstrasse
7, CH-8092 Zuerich, SWITZERLAND

FEATURES

Location/Qualifiers
1. .8121

gene

CDS

/organism="Bradyrhizobium japonicum"
/strain="110spc4"
/db_xref="taxon:375"
complement(1. .440)
/gene="tld"
complement(<1. .440)
/gene="tld"

gene

CDS

/function="putative role in regulation of carbon storage"
/codon_start=1
/transl_table=11
/product="Tld homolog"
/protein_id="CAB56816.1"
/db_xref="GI:6006410"
/translation="MTNEDTSLDLRANLDRDQVHREYARGLAGADGGELEFYSOTE
ALMRDGRKQATYDTSQFGLRAKVDAYASASDPSLPALIRADAVAAVRGYS
GSFSAAPPHTNRLYGDNDPLDAPGFEFTKVKLAELDAYLRDNP"
complement(542. .1075)
/note="ORF177"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56817.1"
/db_xref="GI:6006411"
/translation="WGLSRMARPDSSLNALIGLAALLALPSIAHQGVRSVHG
DWIICDTPPGAQAECCALIOSVAYEDRSNAGLITVLTAKDQSLMVAAPLVGL
GFLSLNGLEGGYDLP"
1523. .2362
/gene="coxB"
1523. .1612
/gene="coxB"
1523. .2362
/gene="coxB"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="cytochrome c oxidase subunit II"
/protein_id="CAB56818.1"
/db_xref="GI:6006412"
/translation="MRMSKRGVRLHGLIAGVLTATGSAFAELCQAPAPWWTLOQ
SGSPVMDNIWVFNFLFVLIATLITFVLTALVVYIKFNANANVPSPRTNHLIEVA
WTLVPLLVIGISVPSFRLFLIEDVPKADLTAKGQWYASYPDNCKFEFDSIM
AOKKORLIGVQENYVYKATRVQVGAUVIAFAFAPGAKIDAIPGLMEYMK
AAKTMFYGCCSELGCKDHAFPIAIRVEDKEFASWETAKKFAAGSGTGTASAG
PTG"

CDS

gene

sig_peptide

CDS

gene

CDS

2451. .4076
/gene="coxA"
2451. .4076
/gene="coxA"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="cytochrome c oxidase subunit I"
/protein_id="CAB56819.1"
/db_xref="GI:6006413"
/translation="MATAAAGHDHGHDEHAPTCMRRYVSTNHKDIGTYLI
FAVAGVGAAMSTAIRAELMTPEVOIFHETHTYNAVETSHGLIMFEVPMAMIGF
LNNVYPLMIGAPMDAPFRNNNISFWLIPASFGLLMSTVEGPGANGVAGMTYV
LSSGHPGPADVAELISLILAGASSILGAINFTTIFNMRACGMLHKMRPLFWSITV
TVFLLLSLPLVAGAILMLLDRNFGTTPADGGCDPVLPHILWFGHEBYVLIIL
PGMFIQIIVSTSRKPVFGYLGMAVAVAGIGIGVYAHMHVTVGSSSAVQAFVA
ATVAVTGVTKIEFMTATMGSSIEFRPAIMAVGFTLFTVGVGTGVVLANGVDR
VLQETVVAHFHYVLSLGNVFAIFAGMYWPEPKMGYVNETLAKAHFWTFIGVNL

gene

CDS

gene

CDS

gene

CDS

gene

CDS

CDS

gene

VEPQHFLGLSCMPRRVYDPAFAGMNIVSSVYISGRVLIPLXCVIADFAKVP
AADNPMGAGATLEWTLPSPPPHQFEVLRVQ"
4137. .5081
/gene="coxE"
4137. .5081
/gene="coxE"
/function="protoheme IX farnesyltransferase"
/codon_start=1
/transl_table=11
/product="putative heme O synthase"
/protein_id="CAB56820.1"
/db_xref="GI:6006414"
/translation="MSYLDONAVDINPERISEAVGDYIALLKPYMSLVIFPALVMA
MARGHPRVLAITSLLCTIAGAGASGALMMALLEGDIDAKMSTRANRPIPRIRTRPA
MTFGMTLAFSSVMTLGIIVMIGALLAFTIFFYVYITMILKRTQNTIYIGAGAG
LPRVAVAAVAGTVDEVELLEAIFPMTPPHMLAIFRSDYARACITPMLPVAGP
DATRLOILLYTIVLIAVAAAPMALGEVAVGVSVLLIAGCMVLATINVMRRRSOS
LRATRKLFASITLYLFALFATLLAEVFPALPMAGGA"
5078. .5251
/gene="coxF"
5078. .5251
/gene="coxF"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56821.1"
/db_xref="GI:6006415"
/translation="MVAADKRPDGIIVLEAOKRSRORSIAIALGLVVLFPVAV
5257. .5901
/gene="coxG"
5257. .5901
/gene="coxG"
/function="biogenesis of cytochrome c oxidase"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="putative heme A synthase or assembly protein"
/protein_id="CAB56822.1"
/db_xref="GI:6006416"
/translation="MDHEPTISROSTARKGVGKGLGPDVLYASICGQVALWAGAS
YAAVEFYNMFCRATGNGTTOVATAPACRILAKIAYRDSNVAAPGLPMKFEPOSZ
IEVNGVATVEYVITNOARTAGCAINAVALTIGVSFOKINCEFTQETMAAGEK
REMPVFEVDPDSIADHENGISTITLSTFFPVKDPVYKPLASGDKRRKGNL"
5981. .6877
/gene="coxC"
5981. .6877
/gene="coxC"
/EC_number="1.9.3.1"
/function="cytochrome c oxidation (respiration)"
/codon_start=1
/transl_table=11
/product="cytochrome c oxidase subunit III"
/protein_id="CAB56823.1"
/db_xref="GI:6006417"
/translation="MATAQKNDHYHLVDPSPMPAVGSLAFIMAVCATIMHINMSA
APVIFGVGVGLYTMASWMDGVLEKADYKDDHRYVOLHNRKMTLFIASEWFEVA
WFWAVNALFEPADAVNATRDVREGCGTAGACAVPGVPRPGIETFDPMHLPILN
TILLISGTTVMANHALLENDROGLKGLITLVVLALEFVQVAYEYSHAASFGN
VVGATPEMATGFGFHVLYGTIFELVLCLEFRAVAGHFTPKOHLGFEFAWYWHFVDVW
LEFLICITYWGHGAETMAHGAH"
6975. .7376
/note="ORF13"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB56824.1"
/db_xref="GI:6006418"
/translation="MNDGAPPKSETVVLQSANRGLACKPCRCGQGLVAGFTLAPS
CDRCGLDAFTDSDGPAIFIMLAGIIVACALIVEVYKQPPRYMLHNVLMPLILAT
TILPLRAKSLILALQLEHNHAAQRGLVDRAK"
7199. .8121
/gene="shb1"


```

HNVRAALDWCGANGDACIGIALAAAPLFIAMSLIECRHMSERALLIDPSTRG
AEBMHIQALG/LTMTFRGQRSGARRIEKPCDGRGARRANOLJOLFMAMHHEBM
GQFDALGYAROSLITVAETLDASLGLASLIGVSLHAGHRDALNMFEMHQBEG
TERSTVHGDFHNRNAGISLARLEIMIGRPGDAROLAROTYEAQDHDHTTICLAI
WLVSIDLWSGELDAEESIDHFIHANSRSGPILSROGKGBELAIRGDADAGVET
IRACIRELHNDAGYELLITTTENIALVHGLLAGOVAOSARLIDAIIRIVEGGDHLVMP
ELLRRRAKAVLISLEPKTGEAETCMOSLELSPGCGESLGAASGD"
2158..2217
misc_feature
/gene="tcrR"
/note="putative"
/complement(4540..5211)
gene
/gene="tcsr"
/complement(4540..5211)
CDS
/gene="tcsr"
/codon_start=1
/transl_table=11
/product="putative two-component system transcriptional
response regulator Tcsr"
/protein_id="AAE78795.1"
/db_xref="GI:8708904"
/translation="MNAPTHLYIADDPLEFADALQAVASVLSARIDEAGSEDT
KLEQTSDDVLDLIDLSMPGISGSLIYRAQYPAIPVIVASDSDATIRSDPG
ASGIPRFGVESYRDLIKVMGDMVVPADPDLSATPDMDRLDRVLTLPQVR
VLMMLSGLLNKQAYELGVSEATIKAHVASILLQKLGESRTOAVIAARIAGQWKQ
CTSTG"
complement(5050..5052)
misc_feature
/gene="tcsr"
/note="putative; phosphorylation site"
5518..7185
/gene="mtbA"
5518..7185
/gene="mtbA"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="integral inner membrane metabolite transport
protein Mtba"
/protein_id="AAE78796.1"
/db_xref="GI:8708905"
/translation="MSTMAVSQAGVGMTRDFVILLASLGTVEFWYDEVLGSLAS
IIGAOFSSAPPTPDIIPALFAAGFLVRPGALVFGIGIVGRKYPITVLING
LSTFIYGLLPNATIGIAPILIALRLAAGLAGEGYGAATYVAHAHNGRKYI
TSFIOTATLGLSLVLTFTSAIGFADPFAMKMRIPFLYSVLLLSIVIRLNL
ESPIPOKMKDEGSKAPITTEPAGNOKKLVLLLGVMGQVWYTGQFVLEFL
OSLIKDYGTANLILMSLLGTFIVGMLSDKIGKPIILGCLLALALFPPIFK
MITSNAPALERAKIEQGGGSGRSKGLRPVQPRYRLHLRHLRTRLPVGVY
FDIIGGRRRREGDQGGSCLRERQGGSCSRRRACRRLSEDRRRHRREDGAVRH
LPSAVAAITGLLIVFTVYGPILAAVLVLPPTLRITYSMSLPYHIGMGFGILL
PATAFAIVASTGDIYGLWYPIVAFISIVIGFLPPEPKVDIKAT"
complement(7350..7823)
gene
/gene="orf157"
/complement(7350..7823)
CDS
complement(7350..7823)
/gene="orf157"
/note="orf157"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="unknown"
/protein_id="AAE78797.1"
/db_xref="GI:8708906"
/translation="MTSARSYVPLGAILIGASHHYLIVAAHGDAGARLHGDPPTAF
IAEVARREGAVTLFRHTSASLTIOENDADPSVLDTLTAISRLAPENAGMTHTEGD
DMRGHKTMLTQTSLOVINGRLALGTWQAIYLIENHARRHREVVLOFTGANO"
complement(7853..8560)
gene
/gene="orf235"
/complement(7853..8560)
CDS
complement(7853..8560)
/gene="orf235"
/note="orf235"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="unknown"

```

```

/protein_id="AAE78798.1"
/db_xref="GI:8708907"
/translation="MPGCRADVPARAGECEBETSPVYNIMAFSGAGRGDGOOR
SCRARIRARRRRRLPDHDADEGRASASTPAPARVLDHOERRRLIGLRACRRARR
RDRPAPDRREAGETARGDTRVCKIDEALPKDPLKAKPKSPRTQDPRAPRPG
RSCDDVPGKGLASADRHGQPRAGRGEDLSLSMRKPGHITRGLSVHSRSLPPEA
VTRIRGNSMPCAIHAH"
8893..9240
gene
/gene="uncI"
/note="orf115"
8893..11493
/gene="Fo AtPase operon"
/note="constitutively expressed"
8893..9240
CDS
/gene="uncI"
/note="orf115; initial gene of the Fo AtPase operon"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="unknown"
/protein_id="AAE78799.1"
/db_xref="GI:8708908"
/translation="METAIIDPRNRVFPNGSEILLISCPNFAATSRPSKPGDGDGR
AARASAMALGFRLSSELVAGVVGAGIGMGPRLISTSPGFIVLLGLFAGVNVV

```

alignment_scores:

Quality:	Ratio:	Length:	Gaps:
Percent Similarity:	83.333	Percent Identity:	40.476

alignment_block:

US-09-712-768-4 x BUJ33883/rev ..

Align seg 1/1 to reverse of: BUJ33883 from: 1 to: 31495

```

2 LeuGlutlleValIrrPThrIleValProvalIleleuValPheIleG1 18
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25233 ATCGAGGTGGCCTGCAGCGCTGTCGCTGATCCTCGTGGCATCTC 25184
18 ylaIheseIeupProvalIleupheIysGlnGlnGluPheProGlnIyA 35
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
25183 GGTGGCGTCTCCGCCCTGCTGTCCTGCGACCTCGACGTCGCCGAAGCGG 25134
35 spIleValIleasnValGluIyArg 43
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25133 ACCTGACCATCAGCGCGCCGCAAG 25108

```

seq_name: gb_sy:SYNBCTCONS

seq_documentation_block:

LOCUS SYNBCTCONS 1260 bp DNA SYN 12-JUN-1993

DEFINITION Bacterial cytochrome oxidase subunit II consensus sequence, 5' end.

ACCESSION L07491

VERSION 107491.1 GI:295264

KEYWORDS cytochrome oxidase II.

SOURCE Artificial gene DNA.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 1260)

Airudozo-peter,R., Escamilla,E., Helman-Meneguz,I.F. and Hernandez,F.

Computer analysis and modelling of a consensus sequence of bacterial cytochrome oxidase subunit II

Unpublished (1992)

FEATURES

location/Qualifiers

1..1260

/organism="synthetic construct"

/db_xref="taxon:32630"

1..1260

/gene="coxIics"

1..>1260

/gene="coxIics"

/EC_number="1.9.3.1"


```

/product="cytochrome oxidase subunit I"
/protein_id="AAC48362.1"
/db_xref="GI:557274"
/translation="MRWIFSTNKHIDIGTLYILFGIWAGLVGALSLLIRAEIGQPGAL
LGDDOYANVIVTAHAEMVIEFLVPMIMIGFGNMVLPLMGVPMAPRINMNSFILL
PPLACILLAGVAGSGAGTGWIVYPPLAGVGHAGSVDLAIFSLIAGVSSILGAVN
FIVTYVMRSEGMOLERLPLFVMSVKITALLIISLTVLAGGITMLITDRNEMTSFED
PAGGDPILYQHLFWRFGHREYVILILPGSGMISHYVMHSSKKEPFGCLYKMYALA
IGLGIIVMAHMEVVGMDVTRAFYFAAMITIAVPGIKTESWLTITGARIGSETP
MMAALGFIPLFTVGGGLGLVLSNSSLDIMLHDSYVVAHFVYLSMGAFARLGAVY
WYDLMTGLSLHERMTKSHFVYVFLGYNLTFEPHFLGLSGMPRYSYDPCYIKMNV
SSIGISLFAVAVLFEMFIVMESIVSGVIMSSLSALEMDSIPDLGFLDKKGLL
FVNILL"
1619. 1682
/feature="codon recognized: GAC"
/product="tRNA-Asp"
/anticodon="(pos:1650..1652,aa:Asp)
2107. 2796
/codon_start=1
/translation="1"
/codon="seq:"tga",aa:(Trp)
/product="cytochrome oxidase subunit II"
/protein_id="AAC48365.1"
/db_xref="GI:557277"
/translation="MAFWSQMGFODGASPIEMOLIFPHDAMILLIITISLGYAVS
LMNSFSLRSLSQSEIYVITLPAVLLFLAPSLDILLDEEPALTIKVVGH
OMWSEYSDFINLEPDSYMSLEDEEGGYRLLEPHRSVPMKTRVRLVTAADVL
HSMVPSLVKADAVPGRNLQSLFFANPPGVFFGQCEICGANSHPILVEYDSSS
FIRKWMNGEA"
2872. 3033
/feature="codon_start=1"
/translation="except=(pos:3001..3003,aa:(Trp))"
/translation="1"
/product="ATPase subunit 8"
/protein_id="AAC48364.1"
/db_xref="GI:3335323"
/translation="MPOLAPMNIIFLLIFWNSVFCVGVVWVWVSSYSLSKSVSW
PEYFKKWM"
3037. 3729
/codon_start=1
/translation="1"
/product="ATPase subunit 6"
/protein_id="AAC48363.1"
/db_xref="GI:557275"
/translation="MMMDIFSSPDNNPTLFSFYLIMVIGLAPVLSQNMWESKSR
FYSMLPKPNMGQVWRAATKNSGFINWISFILLIIVNLSGLPYVPSLSSHS
FAFCGLPEWLSLIFSGDLVSEVSHLPGAGGILNPLVIVETVISIRPTLTSY
RLAANMSAGHITIGLISYLSVGLFCYPIIVLGLLFFVQVFFLEIGVSLQVIFTS
LLVLSYSDHRY"
complement(3758..3822)
/feature="codon recognized: UUC"
/product="tRNA-Phe"
/anticodon="(pos:3789..3791,aa:(Phe)
complement(3821..5536)
/feature="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation="1"
/product="NADH dehydrogenase subunit 5"
/protein_id="AAC48366.1"
/db_xref="GI:557278"
/translation="WKMMFLSTQATVILLSHSLSESTAMTTLTKNKILLFECFYN
FNSVITLPLINMTTALFSAIVCYISASVMIPTKNYSODIFIKRFIVMLVILAS
NLLIFPLATLLGMDGLVSCFVLYONYKSLAGLMTAFMNRIGVAILAIS
VTFSGHNALEFMSTPTMTLISMLLAGTKSQLPFCSMPLPAMAAPPVAVLAV
SEITVAGVFLILIOYPPPLNSHLSTSLIISSTVMAGISANFEDDKIILAST
LSQGVMMLSVSLGLFSLALPHLTHAFKALLPLCAGNIIISHNNDDIKMSHWI
QMPESSTCFNIALNLGCPPEWAGTYSKDVITEKMMFQNNFLISLNFATITLSTAS
SARSLIFWMSKTOSSVSTSDENSLIAPITMLAAGATISGAVFSWLILPSTLPL
ISMANKIATEFATLITGAATYKLFLEKLIKSHSIINPFPMWEMINLSNNMTSSSL
VTGLYTFKSLDQGMETIGEGGLKIMTLITLINOKNOSNLFNIFISLTLMSLTSY
TWMP"
complement(5537..5601)
/feature="codon recognized: CAC"

```

```

/product="tRNA-His"
/anticodon="(pos:5568..5570,aa:(His)
complement(5597..6928)
/feature="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation="except=(pos:complement(6788..6790),aa:(Trp)
/translation="1"
/product="NADH dehydrogenase subunit 4"
/protein_id="AAC48367.1"
/db_xref="GI:3335324"
/translation="WSASTVTKSKSNWIKMSISHPFMTITLNFSPMNPKN
YMFSDNFSVPLMLSCWITGLMSSYKILKNNLVSEFLNWMILNITITMPTOK
SLEFSYIFEEASLITPLILLMGYQPERQAGMYMTITIGALPLILNIFLYSHN
LAAILIMSLPIMPYQAMISFWMFLIIVELFPLIYSFHLMLFKHAEVAGSMI
LAILILKIGAGVINTINNNIPKYNVNLVFTTLTGAGVLSLCLQSDMKALIAV
SSVGHGMLACILSSFERGLKMLIMIAHGLSSGLFMSNMMEYSGSRSLFTRK
GVTLSVPTFSKSFILMCSINMAAPPSINLSAGIMISTVTSKRIILLALMAFVSA
VITLPLFNTQHKISSFVNPNNPFLYVNTFLMLIPADILLSTTTC"
complement(6940..7242)
/codon_start=1
/translation="1"
/product="NADH dehydrogenase subunit 4L"
/protein_id="AAC48368.1"
/db_xref="GI:357280"
/translation="MNFENPLFTGLAUSISALMTICLQKHLINTLALAEILMVS
FILLVLSIMMSNEGIMIFVLLTLAASEASIGIAMLVILIRAHGMDVSSISIRHC"
7247. 7312
/feature="codon recognized: ACA"
/product="tRNA-Thr"
/anticodon="(pos:7275..7277,aa:(Thr)
complement(7312..7378)
/feature="codon recognized: UCA"
/product="tRNA-Ser"
/anticodon="(pos:7348..7350,aa:(Ser)
complement(7381..8520)
/codon_start=1
/translation="except=(pos:complement(8176..8178),aa:(Trp)
/translation="1"
/product="cytochrome b"
/protein_id="AAC48371.1"
/db_xref="GI:3335325"
/translation="MKPRSSHVPVKRIISLIDPSPNLSIMWNEGSLIGLIT
QILTLFLAHETATNIELAFSSVAHICRQVNTGMLIRAHANGSLFPCILFHLGRG
MYGSGFNHITWVGVILFLFMTGATFVLPVGMGSFGATVITNLASAPYIGTT
MYOMINGEFAVDNATLRFESFHLAPFIIAASVAILFHEGTGNNPLGINSSEK
VTIHSFYTLKDIAGFVWVFLFLVFLDPLNADHNEIPNAPVTPHIDPEWFL
FAVAILRSIPNKGIVGGLMSIVLFEVPLINLGRFTISEFVPTQILYMSLVSVL
LTMVGAQVDEPFIILGOTLPFVFLFLMLPDKKMDWTINF"
complement(8513..9019)
/feature="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation="1"
/product="NADH dehydrogenase subunit 6"
/protein_id="AAC48369.1"
/db_xref="GI:557281"
/translation="MPLFLLMSLIFSCARSLPLTPTISGLILLFMSLPISTLIFPG
SSTWFSPIILFIYIGGLVMEAVTALMLPMKRNLIQVFPVSTVFWLMLNLSDF
IGLLLEQTKMFLMNSFFNHLGISLSPENLLIIVSLALILEVFLICVAKICYESNG
PLRPKYA"
9024. 9089
/feature="codon recognized: CCA"
/product="tRNA-Pro"
/anticodon="(pos:9054..9056,aa:(Pro)
complement(9082..10032)
/codon_start=1
/translation="1"
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC48370.1"
/db_xref="GI:557282"
/translation="MPLSFMWVYIAYISILLAAFTLLERKGLGTYORRKGPKYKI
FGLAQLPADALKLFLKEENNLYNSNKLPPYTAIVLSLISLCLMSLYWSDFTLWTFE
GFLFCTSAISVGTMIAGWANSKALRAVAVQIISLVEFLVLLLSAALS
SLNINQSSHQLLYWNSILMTPLIMMIYVCAETNAPDPFAEGSEIVSGNIEYG

```



```

D-loop      /db_xref="taxon:8409"
misc_feature 1. 2425
             /note="TAS (termination associated sequence)"
rep_origin  937..1000
             /note="H-strand origin of replication"
misc_feature 1048..1075
             /note="CSB-1 (conserved sequence block 1)"
misc_feature 1143..1160
             /note="CSB-2 (conserved sequence block 2)"
misc_feature 1264..1282
             /note="CSB-3 (conserved sequence block 3)"
repeat_region 1381..2137
             /rpt_type=tandem
tRNA        2426..2497
             /note="codon recognized: CUA; codon: CUN"
             /product="tRNA-Leu"
             /anticodon="(pos:2458..2460,aa:Leu)
             2500..2569
             /note="codon recognized: ACA"
             /product="tRNA-Thr"
             /anticodon="(pos:2531..2533,aa:Thr)
             complement(2570..2638)
             /note="codon recognized: CCA"
             /product="tRNA-Pro"
             /anticodon="(pos:2606..2608,aa:Pro)
             2640..2709
             /note="codon recognized: UUC"
             /product="tRNA-Phe"
             /anticodon="(pos:2672..2674,aa:Phe)
             2710..3642
             /product="12S ribosomal RNA"
             /note="codon recognized: GUA"
             /product="tRNA-Val"
             /anticodon="(pos:3672..3674,aa:Val)
             3711..5298
             /product="16S ribosomal RNA"
             /note="codon recognized: UUA; codon: UUR"
             /product="tRNA-Leu"
             /anticodon="(pos:5334..5336,aa:Leu)
             5375..6347
             /gene="ND1"
             5375..6347
             /gene="ND1"
             /note="TAA stop codon is completed by the addition of 3' A
             residues to the mRNA."
             /transl_except="(pos:6347,aa:TERM)
             /transl_table=2
             /product="NADH dehydrogenase subunit 1"
             /protein_id="BAB58987.1"
             /db_xref="GI:14278902"
             /translation="MQLMLHLPLLYLAPILLAVALTVERKVLGSAQRKGNVY
             GFGGLPIADGLKLEKEPIRSTSSQILFLATLALAMVMTPEPLVPVSDL
             NSILFLAIFSSIVQVTLIGSGMANSRYALIGARAAGTISYEVTALLICIFL
             AGGTLIAFNESQOOLMEVPLPLAFNMFVSTLAETNRAPDLTEGSESLVSGNVE
             VAGGPFALFETLAEVANILMMNTSLTIFLSLSLSTALMTKASILSLCFLMVR
             ASYPRRIDQMLHVMKNEPLTLTLIIIFHISMPSISLFTPELPKRA"
tRNA        6336..6406
             /note="codon recognized: AUC"
             /product="tRNA-Ile"
             /anticodon="(pos:6367..6369,aa:Ile)
             complement(6407..6477)
             /note="codon recognized: CAA"
             /product="tRNA-Gln"
             /anticodon="(pos:6443..6445,aa:Gln)
             6477..6545
             /note="codon recognized: AUG"
             /product="tRNA-Met"
             /anticodon="(pos:6507..6509,aa:Met)
             6546..7583
             /gene="ND2"
             6546..7583
             /gene="ND2"
             /codon_start=1
             /transl_table=2
             /product="NADH dehydrogenase subunit 2"
             /protein_id="BAB58988.1"
             /db_xref="GI:14278903"
             /translation="MNPPLATLFLFSLAIGTITLSSYHMLIANVGEINTLAIPLM
             TTPHPRATEAKTKYELQATASALLSSLIINAMWGMSINSITDIPMNASIAM
             MKLAGLAPHEMPEVIGISLPGILISIMWKIAPMVILLQSHLININLTIGLTS
             ILVGGGIGGQJQKRTKMAFSSIGHGMMIVYLRAPHLTEFNFTLIEFMTSMELSL
             MISTKMEISISMSKPTPLATLVLLVLSAGLPPLGFAFKLLIVLEIKONATL
             LASLIMLASLALFFYLRILYIVSLPLPNTNPFSSWSKSTPHTAINTALIS
             LPLTPLLLEFI"
tRNA        7582..7651
             /note="codon recognized: UGA"
             /product="tRNA-Trp"
             /anticodon="(pos:7613..7615,aa:Trp)
             complement(7652..7721)
             /note="codon recognized: GCA"
             /product="tRNA-Ala"
             /anticodon="(pos:7688..7690,aa:Ala)
             complement(7722..7794)
             /note="codon recognized: AAC"
             /product="tRNA-Asn"
             /anticodon="(pos:7759..7761,aa:Asn)
             7795..7825
             /note="L-strand origin of replication"
             complement(7822..7888)
             /note="codon recognized: UGC"
             /product="tRNA-Cys"
             /anticodon="(pos:7858..7860,aa:Cys)
             complement(7888..7954)
             /note="codon recognized: UAC"
             /product="tRNA-Tyr"
             /anticodon="(pos:7924..7926,aa:Tyr)
             7959..9497
             /gene="COI"
             7959..9497
             /gene="COI"
             /codon_start=1
             /transl_table=2
             /product="cytochrome oxidase subunit I"
             /protein_id="BAB58989.1"
             /db_xref="GI:14278904"
             /translation="MFRMPFSTNNHKDIGTLYLIFGAMAGVGTALSLIRAEISQPG
             TLICDQIVAVIVTAHAFNMFPMFVMDIILIGFGNNLVPLMIGADPMAPPRMNMSEFM
             LLPSRFFLLASSIVEAGAGTGTVYPRPLAGNALAHGPSVDLAFSLIAGVSLTGA
             INFETTLINNKPTSTIOTYPTLFWMSVLTAVLLLSLYPLANGITMLTLDRNLNTTF
             FDPAGGDPVLYOHLFEWFGHPVEYLLILPFGIISHVAVYSNKRPEYMGVYAM
             LSTIGLFIWAHMHMTTDLNDVTRAVTSATMILAPTVGVKFSMLATMGHGIKWE
             APMLMAGLFIPLFVGGITGIVLANSIDIVLDYVVAHFHYVLSMGAVFAMGF
             VHMPLFTGYLHLELTKRIHVYMETVSNIVLTFEPQHLGAGRPYRSDPDVATLMN
             TVSSVGLSISIVANVMMLFIWEFAKRLFSSELSTNIEMLGLGPPHYHFEESL
             SPLN"
tRNA        complement(9500..9570)
             /note="codon recognized: UCA; codon: UCN"
             /product="tRNA-Ser"
             /anticodon="(pos:9536..9538,aa:Ser)
             9573..9641
             /note="codon recognized: GAC"
             /product="tRNA-Asp"
             /anticodon="(pos:9603..9605,aa:Asp)
             9642..10329
             /gene="COI"
             9642..10329
             /gene="COI"
             /note="TAA stop codon is completed by the addition of 3' A
             residues to the mRNA."
             /codon_start=1
             /transl_except="(pos:10329,aa:TERM)
             /transl_table=2

```

gene

• •
• Thu Dec 27 08:21:50 2001

us-09-712-768-4.rge

US-09-712-768-4 x AAA91490

Align seg 1/1 to: AAA91490 from: 1 to: 132

```

1 ProLeuGIuIleValITpThrIleValIProValIleuValPheI1 17
1 CCGCTGGAAATCGTCTGACGATGTTCCGGTGTGATTCGTGCTCAT
17 eGIyAlaPheSerIeuProValIleuPheIySGInGIuPheProGIug 34
51 CGGCGCTCTCGCTGCGGCTGCTTCAACAGCAGAGTTCGCCGAGG 100
34 IyAspIleValIleAsnValGIuGIyArgSer 44
101 GTGACATCGTCATCAACGTCGAGGCGTGTAGC 132

```

seq_name: /SID58/gcdata/geneseq/geneseqn/NA2001.DAT:AAH70556

seq_documentation_block:

ID AAH70556 standard; cDNA; 496 BP.

XX AAH70556;

DT 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 1830.

KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

OS Homo sapiens.

PN WO200142467-A2.

PD 14-JUN-2001.

PF 08-DEC-2000; 2000WO-US33312.

PR 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 09-JUN-2000; 2000US-0203791.

PR 21-JUL-2000; 2000US-0220114.

PA (MILL-) MILENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Deeds J, Berger A, Zhao X;

DR WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer
 PT and for assessing and detecting compounds for treating the cancer -
 XX
 PS
 XX Claim 1; Page 395; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful: to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.

SQ Sequence 496 BP; 128 A; 142 C; 80 G; 139 T; 7 other;

alignment_scores:

Quality: 64.00

Ratio: 2.207

Percent Similarity: 70.732

Length: 41

Gaps: 0

Percent Identity: 39.024

alignment_block:

US-09-712-768-4 x AAH70556

Align seg 1/1 to: AAH70556 from: 1 to: 496

```

2 LeuGIuIleValITpThrIleValIProValIleuValPheI1eg1 18
344 ATAGAAACCGTTTGAACATCTCTCCCGCATCATCTCATGCTCATTTGG 393
18 yAlaPheSerIeuProValIleuPheIySGInGIuPheProGIugIyA 35
394 CCTCCATCCTACCGCATCTTACATTAACAGCAGAGGTCAACGATCCCT 443
35 spIleValIleAsnValGIuGIy 42
444 CCTTACCATCAATCAATTCGC 466

```

seq_name: /SID58/gcdata/geneseq/geneseqn/NA1996.DAT:AAH73222

seq_documentation_block:

ID AAH73222 standard; DNA; 2167 BP.

XX AAH73222;

DT 15-OCT-1996 (first entry)

DE Macrophage migration inhibitory factor gene.

KW Macrophage migration inhibitory factor; MIF; mitogenesis; cell cycle;

KW cell proliferation; antisense; cancer; tumour; therapy; ss.

OS Homo sapiens.

PN

PD

PF

PR

PR

PR

PR

PA

PI

DR

XX Location/Qualifiers
 FT exon 1..1280
 FT /tag= a
 FT /codon_start= 1173..1175
 FT /tag= b
 FT 1470..1641
 FT /tag= c
 FT 1642..1738
 FT /tag= d
 FT 1739..2167
 FT /tag= e
 FT 1902..1907
 FT /tag= f
 FT polyA_signal

W09615242-A2.

23-MAY-1996.

16-NOV-1995; 95WO-US15000.

16-NOV-1994; 94US-0340826.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Paralkar V, Wistow GJ;

WPI; 1996-259847/26.

P-FSDB: AAR98005.

XX Inhibiting cell proliferation by inhibiting macrophage migration

PT inhibitory factor - partic, using new nucleic acid specific for MIF

XX nucleic acid, used e.g. to treat cancer

PS Disclosure: Page 8-9; 42pp; English.

A DNA sequence (AAH73222) codes for human macrophage migration

CC inhibitory factor (MIF) (AAR98005). MIF binds to the retinoblastoma

CC protein in a cell and is required for entry of the cell into the S

CC phase of the cell cycle. Inhibition of MIF activity arrests cells

CC in the G1 phase, and they cease proliferation. An inhibitory

CC nucleic acid, e.g. an antisense sequence (see also AAT33221) based on
CC the translational start site of mouse MIF mRNA, may be used to
CC control growth of malignant cells, e.g. by direct delivery to a
CC tumour.
XX
SQ Sequence 2167 BP: 392 A; 657 C; 717 G; 401 T; 0 other;

alignment_scores: Length: 33
 Quality: 62.50
 Ratio: 2.976 Gaps: 1
Percent Similarity: 63.636 Percent Identity: 42.424

alignment_block:
US-09-712-768-4 x AAT33222 ..

Align seg 1/1 to: AAT33222 from: 1 to: 2167

```

1  ProLeuGluIleValTTPPhrIleValProValValIleLeuValPheI1 17
   |||||.....:|||||
484 CCACGTGAACCTCTCTGACCTCGCCCTCT.....TCCTT 518
17  egiYAlaPheSerLeuProValLeuPheLysGlnGlnGluPheProGlu 33
   :|||||...:|||||.....:|||||
519 CGGTGGCTCCGACATCCATCTCTTTAAAGCAGGAGGTTTCAGGGA 567

```

seq_name: /SID58/gcgdata/geneseq/geneseq/NA2001.DAT.AAH23190

seq_documentation_block:

ID AAH23190 standard; DNA: 2167 BP.

XX AAH23190;

DT 17-SEP-2001 (first entry)

XX Human macrophage migration inhibitory factor encoding DNA.

DE Macrophage migration inhibitory factor; MIF; antisense; neurological;

KW hyperproliferation; neoplastic; antihormonal; immunosuppressive; human;

KW antiinflammatory; cytostatic; ds.

XX Homo sapiens.

OS WO200153317-A1.

PN 26-JUL-2001.

XX 16-JAN-2001; 2001WO-US01475.

XX 20-JAN-2000; 2000US-0489869.

PR (ISIS-) ISIS PHARM INC.

XX Murray SF, Cowsett LM, Wyatt JR;

PI WPT: 2001-451899/48.

DR P-PSDB; AAB85343.

XX New antisense compound(s) are useful to inhibit a nucleic acid molecule

PT encoding macrophage migration inhibitory factor -

XX Examples: Page 90-92; 105pp; English.

XX The invention relates to antisense oligonucleotides 8-30 nucleotides in

XX length targeted to a nucleic acid molecule encoding macrophage migration

XX inhibitory factor (MIF), where the antisense compound specifically

XX hybridizes with and inhibits the expression of MIF. The antisense

XX nucleotides are useful for the treatment of a disease or condition

XX associated with MIF such as neurological, hormonal, immune, inflammatory

XX or hyperproliferative disorder. The present sequence represents a human

XX MIF encoding DNA.

SQ Sequence 2167 BP: 392 A; 657 C; 717 G; 401 T; 0 other;

alignment_scores: Length: 33
 Quality: 62.50
 Ratio: 2.976 Gaps: 1
Percent Similarity: 63.636 Percent Identity: 42.424

alignment_block:
US-09-712-768-4 x AAH23190 ..

Align seg 1/1 to: AAH23190 from: 1 to: 2167

```

1  ProLeuGluIleValTTPPhrIleValProValValIleLeuValPheI1 17
   |||||.....:|||||
484 CCACGTGAACCTCTCTGACCTCGCCCTCT.....TCCTT 518
17  egiYAlaPheSerLeuProValLeuPheLysGlnGlnGluPheProGlu 33
   :|||||...:|||||.....:|||||
519 CGGTGGCTCCGACATCCATCTCTTTAAAGCAGGAGGTTTCAGGGA 567

```

seq_name: /SID58/gcgdata/geneseq/geneseq/NA1992.DAT.AA20065

seq_documentation_block:

ID AAQ20065 standard; DNA: 6785 BP.

XX AAQ20065;

DT 01-APR-1992 (first entry)

XX Plasmid PA2112.

XX Pneumonia; assay; AIDS; immunosuppressed; ss.

XX Pneumocystis carinii.

XX Location/Qualifiers

EH Key 3..308

FT /tag= a

FT /note= "apocytochrome b"

FT 576..2024

FT /tag= b

FT /note= "NADH dehydrogenase subunit 2"

FT 2035..2370

FT /tag= c

FT /note= "NADH dehydrogenase subunit 3"

FT 2363..3094

FT /tag= d

FT /note= "cytochrome oxidase subunit II"

FT 5236..5760

FT /tag= e

FT /note= "NADH dehydrogenase 6"

FT 5782..6783

FT /tag= f

FT /note= "NADH dehydrogenase 1"

FT 750..766

FT /tag= g

FT /note= "Oligonucleotide 8F"

FT 1285..1271

FT /tag= h

FT /note= "Oligonucleotide 5R"

FT complement (1437..1421)

FT /tag= i

FT /note= "Oligonucleotide 5P"

FT 1954..1970

FT /tag= j

FT /note= "Oligonucleotide 4R"

FT complement (2152..2136)

FT /tag= k

FT /note= "Oligonucleotide 4F"

FT 2607..2623

FT /tag= l

FT /note= "Oligonucleotide 3F"

FT complement (3091..3074)

FT misc_feature

FT misc_feature

FT misc_feature

FT misc_feature

```

FT      /tag= m
FT      /note= "Oligonucleotide 3R"
FT      complement (3739..3723)
FT      /tag= n
FT      /note= "Oligonucleotide 7R"
FT      4354..4338
FT      /tag= o
FT      /note= "Oligonucleotide 2R"
FT      complement (4590..4574)
FT      /tag= p
FT      /note= "Oligonucleotide 2F"
FT      4705..4721
FT      /tag= q
FT      /note= "Oligonucleotide 1R"
FT      5030..5046
FT      /tag= r
FT      /note= "Oligonucleotide 9F"
FT      complement (5315..5299)
FT      /tag= s
FT      /note= "Oligonucleotide 1F"
FT      6017..6033
FT      /tag= t
FT      /note= "Oligonucleotide 6R"
FT      complement (6164..6148)
FT      /tag= u
FT      /note= "Oligonucleotide 6F"
FT      MO9119005-A.
FT      PD 12-DEC-1991.
FT      PF 31-MAY-1991; 91WO-GB00869.
FT      PR 01-JUN-1990; 90GB-0012196.
FT      PA (ISIS-) ISIS INNOVATION LTD.
FT      PI Wakefield AE, Hopkin JM, Moxon ER;
FT      WPI: 1992-007487/01.
FT      DR P-PSDB; AAR20056, AAR21409, AAR21410, AAR21411, AAR21412, AAR21413.
FT      XX New DNA sequences which act as oligo:nucleotide primers - for
FT      PT assaying DNA sample from respiratory secretion of a patient
FT      XX infected with P carinii
FT      PS Claim 2; Fig 3; 42pp; English.
FT      CC This is the complete sequence of the insert of PA2 112, with the
FT      CC oligonucleotide primers indicated. The primers can be used in an
FT      CC assay method for P. carinii. The assay is sensitive and can detect
FT      CC infection. P. carinii is a prime cause of opportunistic pneumonia
FT      CC in patients with AIDS or immunosuppressed on oncology and transplant
FT      CC units. See also AAQ20063-Q20065 and AAQ21396-Q21399.
FT      XX
FT      SO Sequence 6785 BP; 2156 A; 998 C; 1134 G; 2497 T; 0 other;
FT      alignment_scores:
FT      Quality: 62.00 Length: 41
FT      Ratio: 2.000 Gaps: 0
FT      Percent Similarity: 75.610 Percent Identity: 31.707
FT      alignment_block:
FT      US-09-712-768-4 x AAQ20065
FT      Align seg 1/1 to: AAQ20065 from: 1 to: 6785
FT      2 LeuGluIleValTTPprrIleValProValIleLeuValPheIleGI 18
FT      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
FT      2579 ATAGATTGTTGTCAGACGTCCAGCAGCTTACTACTATATAGCATTCG 2628

```

```

18 YAlAPheserLeuProValIleuPheLysGlnGlnGlnIupheProGlnIYA 35
:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2629 TTTTCCAAAGTTTCAAATTAATTTGATTAATGATGATCAAGTATGATTCAT 2678
35 SP1LeValIleAsnValGlnGly 42
:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2679 CCATACACATTTAAACGATAGCT 2701
seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:AA677748
seq_documentation_block:
ID AA677748 standard; DNA; 465 BP.
AC AA677748;
XX
XX 30-JUL-1997 (first entry)
DE
DE H. pylori secreted or periplasmic protein ORF 21687842.aa.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
XX binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
XX duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
XX
XX Helicobacter pylori.
XX
XX Key Location/Qualifiers
XX CDS 1..465
XX
XX /tag= a
XX /transl_except= (pos: 166..168, aa: Xaa)
XX /transl_except= (pos: 169..171, aa: Xaa)
XX /transl_except= (pos: 295..297, aa: Xaa)
XX /transl_except= (pos: 304..307, aa: Xaa)
XX /transl_except= (pos: 319..321, aa: Xaa)
XX /transl_except= (pos: 340..342, aa: Xaa)
XX /transl_except= (pos: 358..360, aa: Xaa)
XX /transl_except= (pos: 376..378, aa: Xaa)
XX /note= "Xaa = unknown"
XX
XX WO940893-A1.
XX
XX PD 19-DEC-1996.
XX
XX PF 06-JUN-1996; 96WO-US09122.
XX
XX PR 01-APR-1996; 96US-0630405.
XX PR 07-JUN-1995; 95US-0487032.
XX
XX PA (ASTR ) ASTRA AB.
XX
XX PI Berglinth OT, Smith D, Mellgaard BJ.
XX DR WPI: 1997-052306/05.
XX DR P-PSDB; AAW20218.
XX
XX Helicobacter pylori nucleic acid sequences and related
XX polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX infection, and to detect Helicobacter
XX
XX Claim 23; Page -: 1481pp; English.
XX
XX This sequence encodes a H. pylori secreted or periplasmic protein.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds,
XX useful as potential H. pylori life cycle activators or inhibitors.
XX The genomic sequence of H. pylori (ATCC 55679) was determined from
XX overlapping contigs generated by mechanically shearing the bacterial
XX DNA. The sequences were analysed by computer evaluation. To
XX identify likely H. pylori antigens for vaccine development, the
XX acid sequences predicted from various ORF were analysed for significant
XX homology to other known or exported membrane proteins. Having identified
XX isolated from H. pylori by PCR amplification for recombinant polypeptide

```

Claim 1; Page 89; 1481pp; English.

This sequence encodes a H. pylori secreted or periplasmic protein. Helicobacter pylori has been strongly linked to chronic gastritis and duodenal ulcer disease. The nucleic acid sequences or inhibitors of are used to evaluate compounds, especially activators or inhibitors of bacterial life cycle, for the ability to bind an H. pylori nucleic acid sequence. The nucleic acid sequences, and corresponding proteins, are also useful for generating vaccines for immunising subjects against H. pylori or for use in detecting the presence of helicobacter species in a sample. Antisense expression of a gene from Helicobacter species used to inhibit genomic DNA was isolated and nebulised to a median size of 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique BstXI-linker adaptors in 100-1000 fold molar excess. These linkers are complementary to the BstXI-cut pmx vectors, while the overhang is not self-complementary. Therefore the linkers will not concatamerise nor will the cut vector re-ligate easily. The linker-adaptor insertions were ligated to each of the 20 pmx vectors to construct a series of shotgun subclone libraries. The purified DNA samples were then sequenced.

ORF/protein reference number for this sequence was obtained Note: The related specification, WO9640893.

Sequence 465 BP; 139 A; 71 C; 108 G; 139 T; 8 other:

	alignment_scores:	Length:
Quality:	60.00	30
Ratio:	2.727	Gaps: 0
Percent Similarity:	73.333	Percent Identity: 40.000

alignment_block:
US-09-712-768-4 x AAT77428 ..

Align seg 1/1 to: AAT77428 from: 1 to: 465

2 LeuGluIleValTrpHrIleValProValValIleLeuValPheIleG1 18
:::||||| ::|||::| |||||
1 ATGAATAATTGTGCCTTATTGTACGGTTAGTTTGTTGTTTTTAGG 50
ATGAATAATTGTGCCTTATTGTACGGTTAGTTTGTTGTTTTTAGG 31
18 yAlapHeSerLeuProValLeuPheLysGlnGlnIupHe 31
||| ::|||::| ::|||::| |||||
51 GCCTGCTGAGAGTTGCCCTGGACTTAATCAAACTCACAAGATTT 90

seq_name: /SINUS8/gcgcdata/geneseq/geneseqn/NA1998.DAT:AAAX30456

seq_documentation_block:
ID AAAX30456 standard: DNA: 504 BP.
AAAX30456:
08-JUN-1999 (first entry)
XX
AC
XX
DT
XX
DE
XX
KM
XX
OS
XX
PN
XX
PD
XX
PE
XX
PF
XX
PR
XX
PR
XX
PA

H. pylori secreted protein ORF 06ep11202_21687842_C3_35.
Vaccine; probe; diagnostic; ORF; cell envelope protein;
secreted protein; cellular protein; ds.
Helicobacter pylori.
W09818323-A1.
07-MAY-1998.
28-OCT-1997; 97WO-US19575.
14-JUL-1997; 97US-0891928.
28-OCT-1996; 96US-0739150.
06-DEC-1996; 96US-0759759.
(ASTR) ASTRA AB.

```

XX XX Alm RA, Smith D;
PI XX
XX DR WPI: 1998-271811/24.
DR P-PDB: AAT10989.
XX XX
PT Helicobacter pylori nucleic acids and proteins - used to develop
PT products for the detection, prevention and treatment of H. pylori
PT infections
XX Claims 3, 4; Page 122; 279pp; English.
PS
XX CC Recombinant or substantially pure preparations of H. pylori polypeptides
CC are disclosed, together with the nucleic acids encoding them. In all,
CC 73 ORFs are shown. The proteins are variously cell envelope proteins,
CC secreted proteins or other cellular proteins. Vaccines containing the
CC 8 nucleic acids or proteins are claimed, as are probes containing at least
CC for treating or reducing the risk of H. pylori infections, and the
CC probes can be used diagnostically for detecting the presence of
CC Helicobacter in a sample. The products are also of use in screening
CC cycle or for compounds having the ability to interfere with the H. pylori life
CC cycle or to inhibit H. pylori infection.
SO Sequence 504 BP; 158 A; 81 C; 116 G; 149 T; 0 other;

alignment_scores:
Quality: 60.00
Ratio: 2.727 Length: 30
Percent Similarity: 73.33 Gaps: 0
alignment_block:
US-09-712-768-4 x AAX30456 ..

Align seg 1/1 to: AAX30456 from: 1 to: 504
2 LeuGluIleValTrpThrIleValProValValIleLeuValPheIleGI 18
1 ATGAATAATTGGTGTCTTTATTTACCGTTAGTTTTTGTTGTTTAGG 50
18 YAlapHeSerLeuProValLeuPheLysGlnGluPhe 31
||||| :::::::::::::::::::: |||||
51 GGCTGTAGACTGCCCTGGAGTTTATCAAACTCAGAATTT 90
seq_name: /SIDSB/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT68220
seq_documentation_block:
ID AAT68220 standard: DNA; 543 BP.
XX
AC AAT68220:
XX
DE 21-JUL-1997 (first entry)
XX
H. pylori secreted or periplasmic protein ORF hp2a1858orf5.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
diagnosis; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..543
FT /tag= "a
FT /note= "no stop codon given"
PD WO9640893-A1.
PN 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US09122.
```

```

XX PR 01-APR-1996; 96US-0630405.
XX PR 07-JUN-1995; 95US-0487032.
XX XX
XX (ASTR ) ASTRA AB.
XX
XX PI Berglindh OT, Smilth D, Mellgaard BL,
XX DR WPI: 1997-052306/05.
XX DR P-PSDB; AAM20967.
XX XX
XX PT Helicobacter pylori nucleic acid sequences and related
XX PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX XX infection, and to detect Helicobacter
XX PS Claim 23; Page 971; 1481pp; English.
XX
XX CC This sequence encodes a H. pylori secreted or periplasmic protein.
XX CC The protein may be used in a vaccine to prevent or treat H. pylori
XX CC infection or to identify H. pylori polypeptide binding compounds,
XX CC useful as potential H. pylori life cycle activators or inhibitors.
XX CC The genomic sequence of H. pylori (ATCC 55679) was determined from
XX CC overlapping contigs generated by mechanically shearing the bacterial
XX CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
XX CC and the predicted coding regions defined by computer evaluation. No
XX CC identity likely H. pylori antigens for vaccine development. The
XX CC acid sequences predicted from various ORF were analysed for significant
XX CC homology to other known or exported membrane proteins. Having identified
XX CC and determined the sequences of interest, particular regions can be
XX CC isolated from H. pylori by PCR amplification for recombinant polypeptide
XX XX production, e.g. in E. coli hosts.
XX
XX SQ Sequence 543 BP; 158 A; 87 C; 127 G; 171 T; 0 other;

alignment_scores:
Quality: 60.00 Length: 30
Ratio: 2.727 Gaps: 0
Percent Similarity: 73.33 Percent Identity: 40.000

alignment_block:
US-09-712-768-4 x AAT68220 ..

Align seg 1/1 to: AAT68220 from: 1 to: 543

2 LeuGluTLeValTrpThrLeValProValAlIleuValPheIleG1 18
:::.....:| | | | | | | | | | | | | | | | | |
22 ATGAAATTGGTAGCTGCTATTCTACGCTGTTGTTCTGTTTTCAGG 71
18 yAlAPheSerLeuProValLeuPheLysGInGInLupHe 31
| | | | | | | | | | | | | | | | | | | | | |
72 GCGCTGACGTTGCTGCTGGAGGTTTATCAACACCAAGAAATT 111

seq_name: /SIDSR/gcgdata/geneseq/geneseqn/MA2001.DAT:AAH54139
seq_documentation_block:
ID AAH54139 Standard; DNA; 3335 BP.
AC AAH54139;
XX
XX 03-SEP-2001 (first entry)
DE
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3503.
XX
XX Staphylococcus epidermidis SRI strain; infection: diagnosis;
XX
XX Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX

```


CC terminates 14 nucleotides upstream from the E2 initiation codon.
CC The E2 and E3 genes have been cloned and can be used for the
CC recombinant production of BCV polypeptides, using e.g. Spodoptera
CC frugiperda Sf9 insect cells as host cells. Glycosylated and
CC non-glycosylated recombinant E2 and E3 are useful as components of
CC vaccines directed toward preventing BCV infection, or reducing the
CC severity of BCV infection, in bovine populations.
XX

SQ Sequence 1305 BP; 331 A; 222 C; 247 G; 505 T; 0 other;

alignment_scores:

Quality:	59.00	Length:	36
Ratio:	2.565	Gaps:	2
Percent Similarity:	63.889	Percent Identity:	38.889

alignment_block:

US-09-712-768-4 x AAT89388 ..

Align seg 1/1 to: AAT89388 from: 1 to: 1305

```
4  ILeValTrpThr...IleValProValValIleLeuVal..... 15
   ::::|||||  ||:::||||  |||||
183 CTTATATGAGACCTTAATCCTGCTGCTGATCTGTGTAATAATATCATCT 232
16  ..... PheIleGlyAlaPheSerLeuProValLeuPheL 27
233 AAGCTGGCAACTCCATTTTAGAGGTTTCACCTTACCGATTTTATAA 282
27  ysgInGln 29
   |||
283 TTACACAG 290
```



```

; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 529
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-111-529

```

```

alignment_scores:
  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

alignment_block:

US-09-712-768-4 x US-09-328-111-529/rev

Align seg 1/1 to reverse of: US-09-328-111-529 from: 1 to: 251

```

2 LeuGIuIleValTrpThrIleValProValValIleLeuValPheIleGI 18
  ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
133 ATGGAACCGCTGACATCTGCTGCGCCGCGCATCTAGTCCTCATGCG 84
  : 18 yAlaPheSerLeuProValLeuPheLysIngIngluPheProGIuGIyA 35
  : | | | | | | | | | | | | | | | | | | | | | | | | | |
83 CCCTCCATCCCTACGATCTTACATACAGACGAGGTCAACGATCCCT 34
  : 35 spIleValIleAsnValGIuGIy 42
  : ::::| | | | | | | | | | | | | | | | | | | | | | | |
33 CCCTTACCATCAATCAATTGCG 11

```

seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-09-385-982-410

seq_documentation_block:

```

; Sequence 410, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-410

```

alignment_scores:

```

  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

alignment_block:

US-09-712-768-4 x US-09-385-982-410

Align seg 1/1 to: US-09-385-982-410 from: 1 to: 353

```

2 LeuGIuIleValTrpThrIleValProValValIleLeuValPheIleGI 18
  ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
218 ATGAACCGCTGACATCTGCTGCGCCGCGCATCTAGTCCTCATGCG 267
  : 18 yAlaPheSerLeuProValLeuPheLysIngIngluPheProGIuGIyA 35
  : | | | | | | | | | | | | | | | | | | | | | | | | | |
268 CCCTCCATCCCTACGATCTTACATACAGACGAGGTCAACGATCCCT 317
  : 35 spIleValIleAsnValGIuGIy 42
  : ::::| | | | | | | | | | | | | | | | | | | | | | | |
318 CCCTTACCATCAATCAATTGCG 340

```

seq_name: /cgn2_6/ptodata/2/ina/5a_COMB.seq:US-08-219-842-2

seq_documentation_block:

```

; Sequence 2, Application US/08219842
; Patent No. 5565323
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; APPLICANT: HerinStadt, Corina
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/219,842
; FILING DATE: 30-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-AG 9504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-219-842-2

```

alignment_scores:

```

  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

alignment_block:

US-09-712-768-4 x US-08-219-842-2

Align seg 1/1 to: US-08-219-842-2 from: 1 to: 754

```

2 LeuGIuIleValTrpThrIleValProValValIleLeuValPheIleGI 18
  ::::| | | | | | | | | | | | | | | | | | | | | | | | | |
251 ATGAACCGCTGACATCTGCTGCGCCGCGCATCTAGTCCTCATGCG 300
  : 18 yAlaPheSerLeuProValLeuPheLysIngIngluPheProGIuGIyA 35
  : | | | | | | | | | | | | | | | | | | | | | | | | | |
301 CCCTCCATCCCTACGATCTTACATACAGACGAGGTCAACGATCCCT 350

```

35 sp1leval1leasnvalglugly 42
: : : : :
351 CCTTACCATCAATCAATTGGC 373

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-451-096-2

seq_documentation_block:
; Sequence 2, Application US/08451096
; Patent No. 5760205
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; APPLICANT: Herrstadt, Corinna
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; TITLE OF INVENTION: for Alzheimer's Disease
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,096
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/219,842
; FILING DATE: 30-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-AG 9504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-451-096-2

alignment_scores:
Quality: 58.00 Length: 41
Ratio: 2.000 Gaps: 0
Percent Similarity: 70.732 Percent Identity: 36.585

alignment_block:
US-09-712-768-4 x US-08-451-096-2 ..

Align seg 1/1 to: US-08-451-096-2 from: 1 to: 754

2 Leugluileval1trphtllevalprovalval1leleuvalphei1eg1 18
: : : : :
251 ATAGAAACGCTGACTGACTCTGCGCCGACATCACTAGTCTCATCGC 300
18 yalapheserleuprovalleuphelsglnclupheproglugly 35
: : : : :
301 CCTCCATCCCTACGATCCTTACATACAGACGAGGTCAACGATCCCT 350
35 sp1leval1leasnvalglugly 42
: : : : :
351 CCTTACCATCAATCAATTGGC 373

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-385-982-539

seq_documentation_block:
; Sequence 539, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 539
; LENGTH: 789
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(789)
; OTHER INFORMATION: n = A,T,C or G
; US-09-385-982-539

alignment_scores:
Quality: 58.00 Length: 41
Ratio: 2.000 Gaps: 0
Percent Similarity: 70.732 Percent Identity: 36.585

alignment_block:
US-09-712-768-4 x US-09-385-982-539/rev ..

Align seg 1/1 to reverse of: US-09-385-982-539 from: 1 to: 789

2 Leugluileval1trphtllevalprovalval1leleuvalphei1eg1 18
: : : : :
133 ATAGAAACGCTGACTGACTCTGCGCCGACATCACTAGTCTCATCGC 84
18 yalapheserleuprovalleuphelsglnclupheproglugly 35
: : : : :
83 CCTCCATCCCTACGATCCTTACATACAGACGAGGTCAACGATCCCT 34
35 sp1leval1leasnvalglugly 42
: : : : :
33 CCTTACCATCAATCAATTGGC 11

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-413-740A-2

seq_documentation_block:
; Sequence 2, Application US/08413740A
; Patent No. 6171859
; GENERAL INFORMATION:
; APPLICANT: HERRNSTADT, CORINNA
; APPLICANT: PARKER, WILLIAM D.
; APPLICANT: DAVIS, ROBERT
; APPLICANT: MILLER, SCOTT W.
; TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
; TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
; NUMBER OF SEQUENCES: 206
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA

```

: ZIP: 20036-5405
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/413,740A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04063
: FILING DATE: 30-MAR-1995
: APPLICATION NUMBER: 08/413,740
: FILING DATE: 30-MAR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Bonham, David B.
: REGISTRATION NUMBER: 34297
: REFERENCE/DOCKET NUMBER: 2105/7
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 429-1776
: TELEFAX: (202) 429-0796
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-413-740A-2

```

```

alignment_scores:
  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

```

alignment_block:
  US-09-712-768-4 x US-08-413-740A-2

```

```

Align seg 1/1 to: US-08-413-740A-2 from: 1 to: 854

```

```

2 LeuGLuIleValTrpThrIleValProValIleLeuValPheIleG1 18
: ::::: ||| ||:::||||:||||:||||:||||:||||:
251 ATGAGAACCCTGACTATCTGCGCCATCATCTAGTCTCATCGC 300
18 yAlaPheSerLeuProValLeuPheLysGlnGlnIupheProGluG1A 35
: ||||| ::::: ||| :
301 CCTCCATCCCTACGATCTTACATACAGAGAGGTCAAGCATCCCT 350
35 splleValIleAsnValGluGly 42
: ::::: |||
351 CCTTACCATCAATCAATTGGC 373

```

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-04063-2

```

```

seq_documentation_block:
: Sequence 2, Application PC/TUS9504063
: GENERAL INFORMATION:
: APPLICANT: HERRNSTADT, CORINNA
: APPLICANT: PARKER, WILLIAM D.
: APPLICANT: DAVIS, ROBERT
: APPLICANT: MILLER, SCOTT W.
: TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
: TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
: NUMBER OF SEQUENCES: 206
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenyon & Kenyon
: STREET: 1025 Connecticut Avenue, N.W.

```

```

: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20036-5405
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04063
: FILING DATE: 30-MAR-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bonham, David B.
: REGISTRATION NUMBER: 34297
: REFERENCE/DOCKET NUMBER: 2105/7
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 429-1776
: TELEFAX: (202) 429-0796
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-04063-2

```

```

alignment_scores:
  Quality: 58.00      Length: 41
  Ratio: 2.000      Gaps: 0
  Percent Similarity: 70.732      Percent Identity: 36.585

```

```

alignment_block:
  US-09-712-768-4 x PCT-US95-04063-2

```

```

Align seg 1/1 to: PCT-US95-04063-2 from: 1 to: 854

```

```

2 LeuGLuIleValTrpThrIleValProValIleLeuValPheIleG1 18
: ::::: ||| ||:::||||:||||:||||:||||:||||:
251 ATGAGAACCCTGACTATCTGCGCCATCATCTAGTCTCATCGC 300
18 yAlaPheSerLeuProValLeuPheLysGlnGlnIupheProGluG1A 35
: ||||| ::::: ||| :
301 CCTCCATCCCTACGATCTTACATACAGAGAGGTCAAGCATCCCT 350
35 splleValIleAsnValGluGly 42
: ::::: |||
351 CCTTACCATCAATCAATTGGC 373

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-415-751-27

```

```

seq_documentation_block:
: Sequence 27, Application US/08415751
: Patent No. 5643772
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUT, JIRI
: TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
: TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
: TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
: TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
: TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PHILLIPS, MOORE, LEMPIO & FINLEY

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: PHILLIPS, MOORE, LEMPJO & FINLEY
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415.751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071,880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/891,301
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolezalova
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Cryptosporidium parvum
US-08-415-751-28

alignment_scores:
Quality: 58.00 Length: 23
Ratio: 3.053 Gaps: 0
Percent Similarity: 82.609 Percent Identity: 34.783

alignment_block:
US-09-712-768-4 x US-08-415-751-28 ..

Align seg 1/1 to: US-08-415-751-28 from: 1 to: 1086

3 GuUleValTRpThIleValProValIleuValPheIleGlyal 19
|||||:||||| :|||:||||| :|||:
726 GAGTGTGTGTGCTTGTCTTGCCCTACATAGTGTGTAAGTGTGTTT 775
19 apheSerLeuProValLeu 25
:|||||:|||||:
776 GCTTGAGCTCCAGTTGTA 794

seq_name: /cgn2_6/ptodata/2/lna/bb_COMB.seq:us-09-097-889-2

seq_documentation_block:
; Sequence 2, Application US/0907889
; Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: Hernstadt, Corina
; APPLICANT: Ghosh, Robert E.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
;

```

COMPUTER: IBM PS/2 Model 50Z or 55SX

```

1 / legYalapheserLeuProValleuPhelys 27
      :::::  :::::
701 GCTACGGATTCAACCTGCGTACGCTGTTTAA 732

```

PRELIMINARY DATA:


```

;
; TELEX: 910/371-7168
;
; INFORMATION FOR SEQ ID NO: 8:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1748 bases
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
US-08-202-056-8

```

```

alignment_scores:
      Quality: 56.00      Length: 28
      Ratio: 2.667      Gaps: 1
Percent Similarity: 75.000      Percent Identity: 42.857

```

```

alignment_block:
US-09-712-768-4 x US-08-202-056-8 ..

```

```

Align seg 1/1 to: US-08-202-056-8 from: 1 to: 1748

```

```

1 ProLeuGIuIleValTrpThr.IIleValProValValIleLeuValPheI 17
  ||| ::::: ||| ||||| ::::: ||| ::::: |||
690 CCGCCCCAGTCCCTTGGCTTCATCGTCCACATGCTGATCATGCTGTCT 739
17 legIyAlapheSerLeuProValIleuPheLys 27
  ::||| ::||| ::||| ::||| ::||| ::|||
740 GCTACGGATTCACCCCTGCTGACGCTGTAAAG 771

```


LOCUS BE518487 705 bp mRNA EST 21-MAY-2001
 DEFINITION EST00063 Atlantic salmon Lambda Zap Express testis cDNA library
 Salmu salar cDNA clone TSM-119 5' similar to cytochrome C oxidase
 chain II, mRNA sequence.
 ACCESSION BE518487
 VERSION BE518487.1 GI:12621507
 KEYWORDS EST.
 SOURCE Atlantic salmon.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
 REFERENCE 1 (bases 1 to 705)
 DAVEY, G.C., CAPLICE, N.C., MARTIN, S.A. and POWELL, R.
 A survey of genes in the Atlantic salmon (Salmo salar) as
 identified by expressed sequence tags
 JOURNAL Gene 263 (1-2), 121-130 (2001)
 MEDLINE 21125223
 COMMENT Contact: Davey GC
 Department of Microbiology
 National University of Ireland Galway (NUIG)
 Galway, Ireland
 Tel: 00353 91 524411 2254
 Fax: 00353 91 525700
 Email: grace.davey@nui.galway.ie
 Insert Length: 719 Std Error: 0.00
 Plate: Testis S+M (E2) row: f column: 08
 Seq primer: M13 reverse primer = caggaacagctatgacc
 High quality sequence stop: 719
 POLYA=yes.

FEATURES
 source location/Qualifiers
 1..705
 /organism="Salmo salar"
 /db_xref="taxon:8030"
 /clone="TSM-119"
 /clone_lib="Atlantic salmon lambda zap Express testis cDNA
 library"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /note="Organ: testis; Vector: Lambda ZAP Express; Site_1:
 EcoRI; Site_2: XhoI; An Atlantic salmon testis cDNA
 library was constructed using the Lambda Zap
 Express/Gigapack cloning kit (Stratagene cloning systems).
 cDNA synthesis was carried out using an oligo-(dT) primer
 for the reverse transcription of 5ug of mRNA and the
 library was constructed by directional cloning EcoRI-XhoI
 based on manufacturers instructions. An insert:vector
 ligation ratio of 1:5 was chosen as most optimum. The
 lambda library was packaged with Gigapack III gold
 packaging extracts and plated on the E. coli cell line
 XL1-blue MRF".
 XL1-blue MRF".

BASE COUNT 205 a 190 c 116 g 192 t 2 others
 ORIGIN

alignment_scores:
 Quality: 79.00 Length: 41
 Ratio: 2.469 Gaps: 0
 Percent Similarity: 78.049 Percent Identity: 39.024

alignment_block:
 US-09-712-768-4 x BE518487 ..

Align seg 1/1 to: BE518487 from: 1 to: 705

```

2 Leugluilevaltrphtlevalprovalvalilleuvalpheilecl 18
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 ATCGAATTCCTTGACCTGCTCCAGCATATCTCATCTTATTGCG 228
18 yalapheserleuprovalleuphelysginglupheprogluglya 35
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 CTTCCCTCTTCGATTCCTTACCTTATAGAGAAATTATATGACCAC 278

```

35 splevalilleasvalgluglya 42
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 279 ACCTTACTTATTAAGCATGGGT 301

seq_name: gb_estl:AW600763

seq_documentation_block:
 LOCUS AW600763 661 bp mRNA EST 30-OCT-2000
 DEFINITION ESTPM2293 Penaeus monodon's total hemocyte cDNA library (#2)
 Penaeus monodon cDNA clone PMC293 5', mRNA sequence.
 ACCESSION AW600763
 VERSION AW600763.1 GI:11037892
 KEYWORDS EST.
 SOURCE black tiger shrimp.
 ORGANISM Penaeus monodon
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
 Penaeidae; Penaeus.
 1 (bases 1 to 661)
 Wongsantichon, J., Boonchoy, C., Udomkit, A., Panyim, S. and
 Sonthayanon, B.
 Expressed Sequence Tags from Black Tiger Prawn's Hemocytes
 Unpublished (2000)
 CONTACT: Sonthayanon, B.
 Prawn Molecular Biology
 Institute of Molecular Biology and Genetics, Mahidol University,
 Salaya campus
 Phuthamonthon 4 Rd., Phuthamonthon District, Nakhon Pathom, 73170
 Thailand
 Tel: 66 2 4419003
 Fax: 66 2 4419006
 Email: scbst@mahidol.ac.th
 Similar to g14582668|emb|X84350.2|MTPNID1 Penaeus notialis
 mitochondrial genome, partial (Acc.no.4582668) cytochrome c oxidase
 subunit II Poly A tract is found after the end of this sequence.
 Seq primer: T3
 POLYA=yes.

FEATURES
 source location/Qualifiers
 1..661
 /organism="Penaeus monodon"
 /db_xref="taxon:6687"
 /clone="PMC293"
 /clone_lib="Penaeus monodon's total hemocyte cDNA library
 (#2)"
 /tissue_type="hemocytes"
 /dev_stage="16-17 g average body weight (farm-raised)"
 /lab_host="E.coli XL1-Blue MRF"
 /note="Vector: LambdaZAP II; Site_1: EcoR I; Site_2: Xho
 I; The library was prepared using protocol given by
 supplier, Stratagene, Inc."

BASE COUNT 209 a 106 c 97 g 248 t 1 others
 ORIGIN

alignment_scores:
 Quality: 77.00 Length: 41
 Ratio: 2.406 Gaps: 0
 Percent Similarity: 78.049 Percent Identity: 36.585

alignment_block:
 US-09-712-768-4 x AW600763 ..

Align seg 1/1 to: AW600763 from: 1 to: 661

```

2 Leugluilevaltrphtlevalprovalvalilleuvalpheilecl 18
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 ATTGAATTTGATGAGACGATCTCCCTCAATTTAATTTTATTGCG 209
18 yalapheserleuprovalleuphelysginglupheprogluglya 35
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 ACTTCCTCTTACGATTCCTTATCTATATAGAGAACTTATATATCCAA 259

```



```

seq.documentat...
LOCUS             467 bp      mRNA                      EST             22-SEP-2000
DEFINITION       AV670026 OLNT cell line cDNA library (OLb) Oryzias latipes cDNA
clone OLB16.03d similar to cytochrome-c oxidase (EC 1.9.3.1) chain II (Atlantic salmon mitochondrion), mRNA sequence.
ACCESSION        AV670026
VERSION          AV670026.1  GI:9935824
KEYWORDS
SOURCE           Japanese medaka.
ORGANISM         Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.
REFERENCE
AUTHORS         Naruse,K., Tanaka,M., Shima,A. and Mitani,H.
TITLE           Medaka EST project in University of Tokyo
JOURNAL         Unpublished (2000)
COMMENT         Contact: Kiyoshi Naruse
                  Department of Biological Sciences
                  Graduate School of Science, University of Tokyo
                  Hongo 7-3-1, Bunkyo-Ku, Tokyo 113-0033, Japan
                  Tel: 81-3-5841-4443
                  Fax: 81-3-5841-4410
                  Email: naruse@biol.s.u-tokyo.ac.jp
                  This clone was isolated from OLNT cell line cDNA library (OLb) 5
                  end sequences.
FEATURES
Source           Location/Qualifiers
                1..467
                /organism="Oryzias latipes"
                /strain="HN1"
                /db_xref="taxon:8090"
                /clone="OLB16.03d"
                /clone_lib="OLNT cell line cDNA library (OLb)"
BASE COUNT       138 a 119 c 77 g 133 t
ORIGIN
alignment_scores:
      Quality: 75.00      Length: 41
      Ratio: 2.344      Gaps: 0
      Percent Similarity: 78.049      Percent Identity: 34.146
alignment_block:
US-09-712-768-4 x AV670026 ..
Align seg 1/1 to: AV670026 from: 1 to: 467
2  LeuGUUleValTTPPhrIIleValPProValIleleuValPheIleGI 18
:::IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
136 ATTGAATATATCTGACCTTCCTGCAGCAATATATCTATATCTGATTCG 185
18 yAlApheserLeuProValLeuPheLysGInGInGluPheProGInGlyA 35
: IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
186 CTTACCCCTCCCTTCGCATCTCTTACCTTATAGATGAATTAATGACCCCC 235
136 ACCTGACACATTAAGCCATAGGC 258
35 spIIleValIIleAsnValGluGly 42
:::IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
236 ACCTGACACATTAAGCCATAGGC 258
seq_name: gb_est1:AV69351
seq.documentat...
LOCUS             469 bp      mRNA                      EST             22-SEP-2000
DEFINITION       AV669351 OLNT cell line cDNA library (OLb) Oryzias latipes cDNA
clone OLB05.00h similar to cytochrome-c oxidase (EC 1.9.3.1) chain II (Atlantic salmon mitochondrion), mRNA sequence.
ACCESSION        AV669351
VERSION          AV669351.1  GI:9934188
KEYWORDS
SOURCE           Japanese medaka.
ORGANISM         Oryzias latipes

```

Thu Dec 27 08:21:57 2001

us-09-712-768-4.1st

Page 5

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:
Acanthomorpha: Acanthopterygii: Percomorpha: Atherinomorpha:
Belontiiformes: Adrianchthyidae: Oryziinae: Oryzias.
1 (bases 1 to 469) Shima, A. and Mitani, H.
Narusue, K., Tanaka, M., Shima, A. and Mitani, H.
Medaka EST Project in University of Tokyo
TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4410
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'
end sequences.
Location/Qualifiers
1..469
/organism="Oryzias latipes"
/strain="HN1"
/db_xref="taxon:8090"
/clone="OLb05.06h"
/clone_lib="OLHNI cell line cDNA library (OLb) 5'
others
136 a 126 c 78 g 127 t
BASE COUNT
ORIGIN
136 a 126 c 78 g 127 t
alignment_scores:
Quality: 75.00 Length: 41
Ratio: 2.344 Gaps: 0
Percent Similarity: 78.049 Percent Identity: 34.146
Percent Identity: 78.049
alignment_block:
US-09-712-768-4 x AV669351 ..
Align seg 1/1 to: AV669351 from: 1 to: 469
seq_name: gb_est1:AV670767
seq_documentation_block:
LOCUS AV670767 OLHNI cell line cDNA library (OLb) Oryzias latipes cDNA
DEFINITION AV670767.05f similar to cytochrome-c oxidase (EC 1.9.3.1) chain
II (Atlantic salmon mitochondrion), mRNA sequence.
ACCESSION AV670767
VERSION AV670767.1 GI:9936565
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4410
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'
end sequences.
Location/Qualifiers
1..502
/organism="Oryzias latipes"
/strain="HN1"
/db_xref="taxon:8090"

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4410
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'
end sequences.
Location/Qualifiers
1..486
/organism="Oryzias latipes"
/strain="HN1"
/db_xref="taxon:8090"
/clone="OLb27.05f"
/clone_lib="OLHNI cell line cDNA library (OLb) 5'
others
141 a 128 c 81 g 136 t
BASE COUNT
ORIGIN
141 a 128 c 81 g 136 t
alignment_scores:
Quality: 75.00 Length: 41
Ratio: 2.344 Gaps: 0
Percent Similarity: 78.049 Percent Identity: 34.146
Percent Identity: 78.049
alignment_block:
US-09-712-768-4 x AV670767 ..
Align seg 1/1 to: AV670767 from: 1 to: 486
seq_name: gb_est1:AV670936
seq_documentation_block:
LOCUS AV670936 OLHNI cell line cDNA library (OLb) Oryzias latipes cDNA
DEFINITION AV670936.07a similar to cytochrome-c oxidase (EC 1.9.3.1) chain
II (Atlantic salmon mitochondrion), mRNA sequence.
ACCESSION AV670936
VERSION AV670936.1 GI:9936734
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4410
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNI cell line cDNA library (OLb) 5'
end sequences.
Location/Qualifiers
1..502
/organism="Oryzias latipes"
/strain="HN1"
/db_xref="taxon:8090"

Thu Dec 27 08:21:57 2001

us-09-712-768-4.fst

BASE COUNT 148 a 130 c 82 g 142 t
ORIGIN

alignment_scores:

Quality: 75.00 Length: 41
Ratio: 2.344 Gaps: 0
Percent Similarity: 78.049 Percent Identity: 34.146
US-09-712-768-4 x AV670936 ..
Align seg 1/1 to: AV670936 from: 1 to: 502

2 LeuGLuIleValTrpThrIleValProValIleLeuValPheIleGI 18
168 ATTGAATATATCTGCACCTCTCCAGCAATATATCTGATGCC 217
18 yAlaPheSerLeuProValLeuPheIleValGluGly 35
218 CTTACCCCTCCCTGCATCTTACCTTATAGATGAATATATGACCC 267
35 spIleValIleAsnValGluGly 42
268 ACCGACAAATTAAGCCATAGCC 290
seq_name: gb_est1:AV669884

seq_documentation_block:
LOCUS AV669884 OLHNT cell line cDNA library (OLB) EST
DEFINITION

clone OLB3.03c similar to cytochrome-c oxidase (EC 1.9.3.1) chain
VERSION AV669884
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei;
Belontiiformes; Acanthopterygii; Percomorpha; Atherinomorpha;
Nuruse, K., Tanaka, M., Shima, A. and Mitani, H.
Medaka EST Project in University of Tokyo.
TITLE Unpublished (2000)
JOURNAL Contact: Kiyoshi Naruse
COMMENT Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNT cell line cDNA library (OLB) and sequences.

REFERENCE 1 (bases 1 to 536)
AUTHORS Naruse, K., Tanaka, M., Shima, A. and Mitani, H.
TITLE Unpublished (2000)
JOURNAL Contact: Kiyoshi Naruse
COMMENT Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNT cell line cDNA library (OLB) and sequences.

FEATURES
Source Location/Qualifiers
1..536
/organism="Oryzias latipes"
/strain="HNT"
/db_xref="taxon:8090"
/clone="OLB13.03c"
/clone_lib="OLHNT cell line cDNA library (OLB)"
BASE COUNT 153 a 139 c 91 g 153 t
ORIGIN

alignment_scores:

Quality: 75.00 Length: 41
Ratio: 2.344 Gaps: 0
Percent Similarity: 78.049 Percent Identity: 34.146

alignment_block:

US-09-712-768-4 x AV669884 ..

Align seg 1/1 to: AV669884 from: 1 to: 536

2 LeuGLuIleValTrpThrIleValProValIleLeuValPheIleGI 18
120 ATTGAATATATCTGCACCTCTCCAGCAATATATCTGATGCC 169
18 yAlaPheSerLeuProValLeuPheIleValGluGly 35
170 CTTACCCCTCCCTGCATCTTACCTTATAGATGAATATATGACCC 219
35 spIleValIleAsnValGluGly 42
220 ACCGACAAATTAAGCCATAGCC 242
seq_name: gb_est1:AV670737

seq_documentation_block:
LOCUS AV670737
DEFINITION

clone OLB26.12b similar to cytochrome-c oxidase (EC 1.9.3.1) chain
VERSION AV670737
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei;
Belontiiformes; Acanthopterygii; Percomorpha; Atherinomorpha;
Nuruse, K., Tanaka, M., Shima, A. and Mitani, H.
Medaka EST Project in University of Tokyo.
TITLE Unpublished (2000)
JOURNAL Contact: Kiyoshi Naruse
COMMENT Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNT cell line cDNA library (OLB) and sequences.

REFERENCE 1 (bases 1 to 555)
AUTHORS Naruse, K., Tanaka, M., Shima, A. and Mitani, H.
TITLE Unpublished (2000)
JOURNAL Contact: Kiyoshi Naruse
COMMENT Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from OLHNT cell line cDNA library (OLB) and sequences.

FEATURES
Source Location/Qualifiers
1..555
/organism="Oryzias latipes"
/strain="HNT"
/db_xref="taxon:8090"
/clone="OLB26.12b"
/clone_lib="OLHNT cell line cDNA library (OLB)"
BASE COUNT 155 a 140 c 99 g 161 t
ORIGIN

alignment_scores:

Quality: 75.00 Length: 41
Ratio: 2.344 Gaps: 0
Percent Similarity: 78.049 Percent Identity: 34.146
US-09-712-768-4 x AV670737 ..

Align seg 1/1 to: AV670737 from: 1 to: 555

2 LeuGLuIleValTrpThrIleValProValIleLeuValPheIleGI 18
96 ATTGAATATATCTGCACCTCTCCAGCAATATATCTGATGCC 145
18 yAlaPheSerLeuProValLeuPheIleValGluGly 35
||||| ..|||

AUTHORS
 TITLE
 JOURNAL
 COMMENT

Rojinakorn, J., Hirono, I., Aoki, T., Itami, T. and Takahashi, Y.
 Gene expression in non-infected and viral infected hemocytes of
 kuruma prawn (Penaeus japonicus)
 Unpublished (2001)
 Contact: Takashi Aoki
 Aquatic Biosciences
 Tokyo University of Fisheries
 Konan 4-5-7, Tokyo, Minato-ku 108-8477, Japan
 Tel: 81-3-5463-0689
 Fax: 81-3-5463-0690
 Email: aoki@tokyo-u-fish.ac.jp.

FEATURES

source
 1 .579
 location/Qualifiers
 /organism="Marsupenaeus japonicus"
 /db_xref="taxon:27405"
 /clone_id="PjH325"
 /clone_lib="Marsupenaeus japonicus adult"
 /cell_type="hemocytes"
 /dev_stage="adult"
 /note="Vector: lambda ZAP II"
 BASE COUNT 202 a 95 c 121 g 160 t 1 others
 ORIGIN

alignment_scores:
 Quality: 75.00 Length: 41
 Ratio: 2.419 Gaps: 0
 Percent Similarity: 75.610 Percent Identity: 36.585

alignment_block:
 US-09-712-768-4 x AU175272/rev ..

Align seg 1/1 to reverse of: AU175272 from: 1 to: 579

```

2 LeuGLIITLeValTrpThrIleValProValValIleLeuValPheIleG1 18
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
479 ATTGAATGTGTGACAGTACTTCGCGCCCTTATCTTATCTCATCTCGC 430
18 yAlaPheSerLeuProValLeuPheLysGlnGlnGluPheProGluG1 35
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
429 CCTACTTCTTACGATTTATTATTACTAGATGAGTTCATATCCNA 380
35 spIleValIleAsnValGluG1 42
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
379 GAGTTACATTTGAGACATTTGCT 357
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:01:48 ; Search time 39.94 Seconds
(without alignments)
81.603 Million cell updates/sec

Title: US-09-712-768-4

Sequence: 1 PLEIWTIVPVILVIFGAF.....LFKQEPFEGDIVIVEGRS 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_1101:*

```

1: /SID88/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID88/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID88/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID88/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID88/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID88/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID88/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID88/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID88/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID88/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID88/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID88/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID88/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID88/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID88/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID88/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID88/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID88/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID88/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID88/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID88/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID88/gcgdata/geneseq/geneseqp/AA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	222	100.0	44	AAV97751	G. oxydans cytochr
2	75	33.8	129	AAAB39150	Human secreted pro
3	62	27.9	243	AAAB21411	Cytochrome oxidase
4	60	27.0	154	AAW20218	H. pylori secreted
5	60	27.0	154	AAW24610	H. pylori secreted
6	60	27.0	167	AAV10989	H. pylori ORF 06p
7	59	27.0	181	AAW20967	H. pylori secreted
8	59	26.6	20	AAV97755	G. oxydans cytochr
9	59	26.6	55	AAAG03705	Human secreted prote
10	59	26.6	197	AAAG90866	C glutaminc prote
11	56.5	25.5	593	AAE05103	Drosophila melanog

12	56	25.2	359	22	AAAG2151
13	56	25.2	359	22	AAAB79573
14	55	24.8	132	21	AAAB39151
15	54	24.3	1019	19	AAW71527
16	54	24.3	1020	19	AAW98661
17	54	24.3	1020	22	AAAB46344
18	53.5	24.1	363	18	AAW08437
19	53.5	24.1	363	18	AAAG81185
20	53	23.9	83	22	AAW24378
21	53	23.9	391	21	AAAG06185
22	53	23.9	391	21	AAAG49300
23	52.5	23.6	113	21	AAV87356
24	52.5	23.6	314	19	AAW37817
25	52.5	23.6	314	20	AAW88528
26	52	23.4	40	22	AAAG75937
27	52	23.4	40	22	AAAG75937
28	52	23.4	1291	16	AAW75201
29	52	23.4	1816	21	AAV95440
30	51.5	23.2	259	20	AAV35625
31	51.5	23.2	691	11	AAAR05234
32	51.5	23.2	1169	16	AAW76544
33	51	23.0	193	21	AAAG37400
34	51	23.0	193	21	AAAG37422
35	51	23.0	212	21	AAAG37398
36	51	23.0	212	21	AAAG37420
37	51	23.0	385	21	AAAG06186
38	51	23.0	385	21	AAAG49301
39	51	23.0	407	19	AAW80618
40	51	23.0	407	21	AAW81718
41	51	23.0	407	22	AAW01042
42	51	23.0	437	21	AAAG13750
43	51	23.0	437	21	AAAG20978
44	51	23.0	438	21	AAAG38871
45	51	23.0	447	21	AAAG13749

ALIGNMENTS

RESULT 1					
ID	AAV97751	standard; Protein; 44 AA.			
XX	AAV97751;				
AC	AAV97751;				
XX	06-AUG-2001	(first entry)			
DE	G. oxydans cytochrome C oxidase (COII) protein sequence fragment.				
XX					
XX	Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;				
KW	oxidative fermentation; electron transfer; respiratory chain; L-sorbose;				
KW	2-Keto-L-gluconic acid; 2KGa production; aldehyde production;				
KW	carboxylic acid production; ketone production.				
XX					
OS	Glucanobacter oxydans.				
XX					
PN	EP1103603-A2.				
XX					
PD	30-MAY-2001.				
XX					
PF	14-NOV-2000; 2000EP-0124785.				
XX					
PR	17-NOV-1999; 99EP-0122842.				
XX					
PA	(HOFF) HOFFMANN LA ROCHE & CO AG F.				
XX					
PI	Asakura A, Hoshino T, Shinjoh M;				
XX					
DR	WPI; 2001-357953/38.				
XX					
DR	N-PSDB; AAA91490.				
XX					
PT	New cytochrome c oxidase complex having cytochrome c oxidase activity from Glucanobacter oxydans DSM 4025, useful in mediating electron				

C glutaminc prote
Corynebacterium gl
Human secreted pro
Helicobacter polyp
H. pylori GHPD 470
H. pylori HPS132 p
MWB2-92 full lengt
Mycobacterium tube
Rat EST encoded pr
Arabidopsis thalia
Arabidopsis thalia
Human signal pepti
Human secreted apo
Human Fritzed-rel
Human colon cancer
Human colon cancer
Tyrosine phosphat
Caenorhabditis ele
Chlamydia pneumoni
Sequence of yeast
Mitochondrial cyto
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
S. pneumoniae prot
Streptococcus pneu
CFE 45 protein seq
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia

PT transfer in respiratory chain or producing 2-keto-L-gulononic acid from
 PT L-sorbose or D-sorbitol
 XX
 XX
 PS Claim 6; Page 24; 42pp; English.

CC This sequence represents a fragment of the gluconobacter oxydans
 CC cytochrome C oxidase, COI1. The invention relates to a cytochrome c
 CC oxidase COI complex. The COI complex is useful in improving oxidative
 CC fermentation and is an essential component mediating electron transfer
 CC in the respiratory chain. The recombinant microorganism and the
 CC cytochrome c oxidase may be used in the genetic preparation of a
 CC recombinant COI complex and in the production of 2-Keto-L-gulononic acid
 CC (2KA) from L-sorbose or D-sorbitol in a culture medium. The COI is also
 CC useful as a terminal oxidase, oxidising cytochrome c, an electron
 CC acceptor from an enzyme belonging to dehydrogenase c, for the production of
 CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,
 CC especially the production of 2KA from L-sorbose or D-sorbitol.

XX
 SQ Sequence 44 AA:

Query Match 100.0%; Score 222; DB 22; Length 44;
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLEIWTIVPVVILVFGAFLPVLFKQEPFEGDIVINVEGRS 44
 1 PLEIWTIVPVVILVFGAFLPVLFKQEPFEGDIVINVEGRS 44

DB 1 PLEIWTIVPVVILVFGAFLPVLFKQEPFEGDIVINVEGRS 44

RESULT 2

AAB39150
 ID AAB39150 standard; Protein; 129 AA.

XX AAB39150;

DT 02-FEB-2001 (first entry)

DE Human secreted protein BLAST search protein SEQ ID NO: 117.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

OS WO200058513-A1.

XX 05-OCT-2000.

PD 22-MAR-2000; 2000WO-US07506.

XX 26-MAR-1999; 99US-0126505.

PR 17-DEC-1999; 99US-0172412.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-594649/56.

XX Forty-nine polynucleotide sequences, and their encoded secreted
 PT polypeptides, used in the treatment and diagnosis of cancers,
 PT autoimmune disorders, and skin disorders -

PS Disclosure; Page 394; 413pp; English.

XX The invention relates to the isolation of genes AAC73865-C73913 encoding
 CC the human secreted proteins AAB39093-839141. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene isolated

CC in the present invention. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
 CC as myocardial ischaemias; (d) wound healing; (e) neurological diseases
 CC as viral, bacterial, fungal and parasitic infections.

XX
 SQ Sequence 129 AA:

Query Match 33.8%; Score 75; DB 21; Length 129;
 Best Local Similarity 39.0%; Pred. No. 0.0022;
 Matches 16; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

QY 2 LEIWTIVPVVILVFGAFLPVLFKQEPFEGDIVINVEG 42
 36 metwtllpalllvllalplrllymtdevnpsltksig 76

DB 36 metwtllpalllvllalplrllymtdevnpsltksig 76

RESULT 3

AAR21411
 ID AAR21411 standard; Protein; 243 AA.

XX AAR21411;

DT 01-APR-1992 (first entry)

DE Cytochrome oxidase subunit II.

XX Pneumonia; assay; AIDS; immunosuppressed.

OS Pneumocystis carinii.

XX WO9119005-A.

XX 12-DEC-1991.

XX 31-MAY-1991; 91WO-GB00869.

PR 01-JUN-1990; 90GB-0012196.

XX (ISIS-) ISIS INNOVATION LTD.

PI Wakefield AE, Hopkin JM, Moxon ER;

XX WPI; 1992-007487/01.

DR N-PSDB; AAO20065.

XX New DNA sequences which act as oligo:nucleotide primers - for
 PT assaying DNA sample from respiratory secretion of a patient
 PT infected with P carinii

PS Claim 3; Fig 3; 42pp; English.

XX The amino acid sequence is that of P. carinii cytochrome oxidase
 CC subunit II which was translated from DNA from plasmid pA112. See
 CC also AAR21409-R21413 and AAR20056.

XX
 SQ Sequence 243 AA:

Query Match 27.9%; Score 62; DB 13; Length 243;
 Best Local Similarity 31.7%; Pred. No. 0.39;
 Matches 13; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFIGAFLPVLFEKQEPPEGDIVINVG 42
 Db 73 lefwtvspalllatafpstkilylmdeviopsitikaig 113

RESULT 4
 AAW20218
 ID AAW20218 standard; Protein: 154 AA.
 AC AAW20218;
 XX
 XX 30-JUL-1997 (first entry)

DE H. pylori secreted or periplasmic protein 21687842.aa.
 XX
 XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX Helicobacter pylori.
 OS

XX Key Location/Qualifiers
 FH Misc-difference 56 /label= "unknown"
 FT /note= "encoded by RAR"
 FT Misc-difference 57 /label= "unknown"
 FT /note= "encoded by GRC"
 FT Misc-difference 99 /label= "unknown"
 FT /note= "encoded by TRP"
 FT Misc-difference 102 /label= "unknown"
 FT /note= "encoded by TRP"
 FT Misc-difference 107 /label= "unknown"
 FT /note= "encoded by TMC"
 FT Misc-difference 114 /label= "unknown"
 FT /note= "encoded by ART"
 FT Misc-difference 120 /label= "unknown"
 FT /note= "encoded by RAR"
 FT Misc-difference 126 /label= "unknown"
 FT /note= "encoded by GRC"

XX WO9640893-A1.
 PN
 XX 19-DEC-1996.
 PD
 XX 06-JUN-1996; 96WO-US09122.
 PF
 XX 01-APR-1996; 96US-0630405.
 PR
 PR 07-JUN-1995; 95US-0487032.
 XX (ASTR) ASTRA AB.
 PA
 XX Berglindh OT, Smith D, Møllgaard BL;
 PI
 DR WPI: 1997-052306/05.
 N-PSDB: AAT67749.

XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 72; Page 431; 1481pp; English.

CC This sequence is a H. pylori secreted or periplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. No
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC
 SQ Sequence 154 AA;

Query Match 27.0%; Score 60; DB 18; Length 154;
 Best Local Similarity 40.0%; Pred. No. 0.45; 10; Indels 0; Gaps 0;
 Matches 12; Conservative 8; Mismatches 10;
 OY 2 LEIWTIVPVILVFIGAFLPVLFEKQEP 31
 Db 1 mkvslvalvfcctlgavelpgyqtgef 30

RESULT 5
 AAW24610
 ID AAW24610 standard; Protein: 154 AA.
 AC AAW24610;
 XX
 XX 11-AUG-1997 (first entry)

DE H. pylori secreted or periplasmic protein 21687842.aa.
 XX
 XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
 KW detection; antisense; inhibition.
 XX
 OS Helicobacter pylori.

XX Key Location/Qualifiers
 FH Misc-difference 56 /label= "unknown"
 FT /note= "encoded by RAR"
 FT Misc-difference 57 /label= "unknown"
 FT /note= "encoded by GRC"
 FT Misc-difference 99 /label= "unknown"
 FT /note= "encoded by TRP"
 FT Misc-difference 102 /label= "unknown"
 FT /note= "encoded by TRP"
 FT Misc-difference 107 /label= "unknown"
 FT /note= "encoded by TMC"
 FT Misc-difference 114 /label= "unknown"
 FT /note= "encoded by ART"
 FT Misc-difference 120 /label= "unknown"
 FT /note= "encoded by RAR"
 FT Misc-difference 126 /label= "unknown"
 FT /note= "encoded by GRC"

XX WO9719098-A1.
 PN
 XX 29-MAY-1997.
 PD
 XX 15-NOV-1996; 96WO-US18542.
 PF
 XX 17-NOV-1995; 95US-0561469.
 PR
 PR (ASTR) ASTRA AB.
 PA
 XX Smith DH;

XX WPI: 1997-298052/27.
 DR N-PSDB: AAT77428.

CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.

CC Sequence 181 AA:

Query Match 27.0%; Score 60; DB 18; Length 181;
Best Local Similarity 40.0%; Pred. No. 0.54;
Matches 12; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGASLPLVLFKQDEF 31
DB 8 mklvalvalvfccflgavelpgytqet 37

RESULT 8

AAV97755 AAV97755 standard; Peptide: 20 AA.

AC AAV97755;

DT 06-AUG-2001 (first entry)

DE G. oxydans cytochrome C oxidase (COII) protein sequence fragment.

XX Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;
XX oxidative fermentation; electron transfer; respiratory chain; L-sorbose;
XX 2-keto-L-gluconic acid; 2KGA production; aldehyde production;
XX carboxylic acid production; ketone production.

OS Gluconobacter oxydans.

PN EPI103603-A2.

PD 30-MAY-2001.

PF 14-NOV-2000; 2000EP-0124785.

PR 17-NOV-1999; 99EP-0122842.

PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

PI Asakura A, Hoshino T, Shinjoh M;

DR WPI; 2001-357953/38.

XX New cytochrome c oxidase complex having cytochrome c oxidase activity
PT from Gluconobacter oxydans DSM 4025, useful in mediating electron
PT transfer in respiratory chain or producing 2-keto-L-gluconic acid from
PT L-sorbose or D-sorbitol

PS Example 2; Page 9; 42pp; English.

XX This sequence represents a fragment of the Gluconobacter oxydans
CC cytochrome C oxidase, COII. The invention relates to a cytochrome c
CC oxidase COI complex. The COI complex is useful in improving oxidative
CC fermentation and is an essential component mediating electron transfer
CC in the respiratory chain. The recombinant microorganism and the
CC cytochrome c oxidase may be used in the genetic preparation of a
CC recombinant COI complex and in the production of 2-keto-L-gluconic acid
CC (2KGA) from L-sorbose or D-sorbitol in a culture medium. The COI is also
CC useful as a terminal oxidase, oxidising cytochrome c, an electron
CC acceptor from an enzyme belonging to dehydrogenase for the production of
CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,
CC especially the production of 2KGA from L-sorbose or D-sorbitol.

SO Sequence 20 AA:

Query Match 26.6%; Score 59; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PLEWTIVPV 11
DB 10 plewtivpv 20

RESULT 9

AAAG03705 AAAG03705 standard; Protein: 55 AA.

AC AAAG03705;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 7786.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.

OS Homo sapiens.

PN EPI103401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GENEST) GENEST.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

DR N-PSDB; AAC03711.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 13; SEQ ID 7786; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

PS Sequence 55 AA:

Query Match 26.6%; Score 59; DB 21; Length 55;
Best Local Similarity 42.4%; Pred. No. 0.19; Mismatches 15; Indels 0; Gaps 0;
Matches 14; Conservative 4;

OY 7 TTPVVIIVIFGASLPLVLFKQDEPGDIVIN 39
DB 20 tsvpvflvcvvrkpvlysqststqstvin 52

RESULT 10

AAG90866

ID AAG90866 standard; Protein: 197 AA.

XX AAG90866;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 4620.

XX Corynebacterium: amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.

XX Corynebacterium glutamicum.

XX EPI108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOCYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

XX N-PSDB: AAH66085.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

XX mutation point of a gene, measuring expression of a gene, analysing

XX expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 4620; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

XX sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These

XX are useful for identifying the mutation point of a gene derived from a

XX mutant of corynebacterium bacterium, measuring expression amount and

XX analysing the expression profile or expression pattern of a gene derived

XX from Corynebacterium bacterium, and identifying a homologue of a gene derived

XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

XX particularly L-lysine. The present sequence is a protein described

XX in the exemplification of the invention.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from the

XX European Patent Office.

XX Sequence 197 AA:

XX Query Match 26.6%; Score 59; DB 22; Length 197;

XX Best Local Similarity 33.3%; Pred. No. 0.84;

XX Matches 19; Conservative 9; Mismatches 11; Indels 18; Gaps 3;

XX QY 4 IWTIVPVVILVFITGAFS---LPV-----LEKQOE-----FPEGDIVINVEG 42

XX DB 120 IYIGTAPAILLVfagafnglvipvgftlmlyvalfrgdkllkgykypwllvvgvlg 176

RESULT 11

AAE05103

ID AAE05103 standard; Protein: 593 AA.

XX AAE05103;

XX

DT 18-SEP-2001 (first entry)

XX Drosophila melanogaster dmsNF2.

XX Fruit fly; invertebrate symporter cell surface receptor; dmsNF2;

XX sodium/neurotransmitter family; SNF; biopesticide; therapy.

XX Drosophila melanogaster.

XX WO200149848-A2.

XX 12-JUL-2001.

XX 28-DEC-2000; 2000WO-US35551.

XX 30-DEC-1999; 99US-0173929.

XX 15-MAR-2000; 2000US-0189399.

XX 23-MAR-2000; 2000US-0191686.

XX 23-MAR-2000; 2000US-0191687.

XX 23-MAR-2000; 2000US-0191688.

XX 23-MAR-2000; 2000US-0191695.

XX (GENO-) GENOPTERA LLC.

XX Kellerman KA, Keegan KP, Ebens AJ, Torpey J;

XX WPI: 2001-441879/47.

XX N-PSDB: AAD09683.

XX Novel invertebrate symporter cell surface receptor proteins and nucleic

XX acid encoding the protein useful as pesticide or drug target and to

XX identify compounds that have utility as therapeutics or pesticides

XX Claim 18; Page 65-66; 71pp; English.

XX The invention relates to invertebrate symporter cell surface receptors

XX of the sodium/neurotransmitter family (SNF) and nucleic acid molecules

XX encoding such receptors. The SNF protein is useful for detecting a

XX candidate compound especially a putative pesticide or pharmaceutical

XX agent that interacts with an invertebrate symporter cell surface

XX receptor protein or its fragment. Insect or worm genetically modified

XX symporter cell surface receptor protein activity, by detecting the

XX phenotype caused by the expression or mis-expression of the protein in

XX the animal. Nucleic acids encoding the invertebrate receptor protein or

XX their fragments are useful as biopesticides. SNF nucleic acids are

XX useful for generating mutant phenotypes in an animal model or living

XX cells that are used to study the regulation of genes encoding the

XX modified organisms or cells are useful in screening assays to identify

XX CC pesticides or therapeutics and thus are useful in the identification of

XX CC new drug targets, therapeutic agents, diagnostics and prognostics

XX CC useful in treatment of disorders associated with ion channels. The

XX CC nucleic acid molecules are also useful as hybridisation probes. The

XX CC The present sequence is Drosophila melanogaster (dm) SNF homologue

XX CC which is referred as dmsNF2.

XX Sequence 593 AA:

XX Query Match 25.5%; Score 56.5; DB 22; Length 593;

XX Best Local Similarity 34.9%; Pred. No. 7.2;

XX Matches 15; Conservative 9; Mismatches 12; Indels 7; Gaps 2;

XX QY 4 IWTIVPVVILVFITGAFSL---PVLFEQOEFPFEGDIVINVEG 42

XX DB 472 IOWSIVPLVlvmllvlyslldmrlpsyngefp---lvrvvg 511

RESULT 12

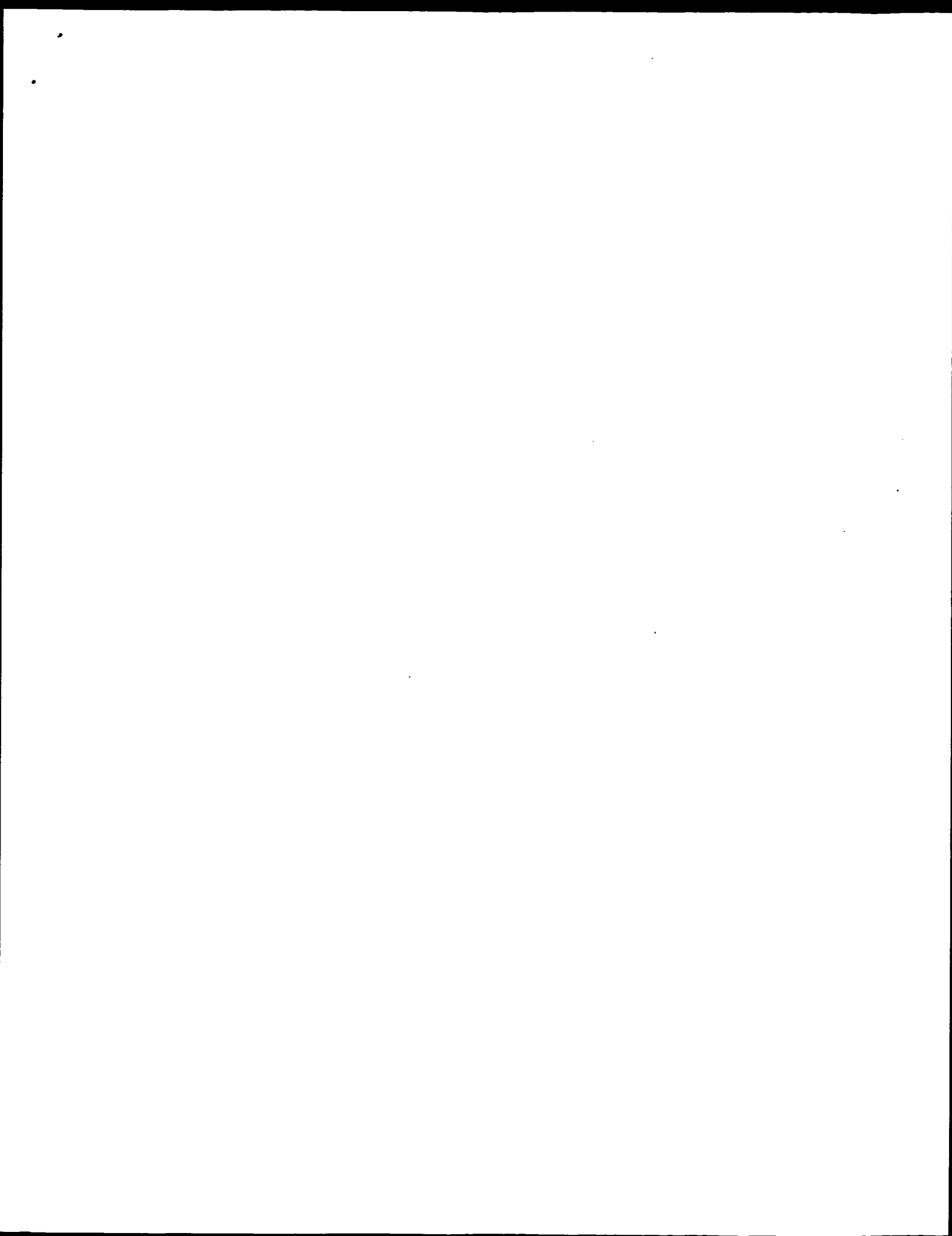
AAG92151

ID AAG92151 standard; Protein: 359 AA.

XX

AC AAG92151;
 XX
 XX 26-SEP-2001 (first entry)
 XX
 XX C glutamicum protein fragment SEQ ID NO: 5905.
 DE
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 XX Corynebacterium glutamicum.
 OS
 XX EPI108790-A2.
 XX
 XX 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 PF
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI: 2001-376931/40.
 DR N-PSDB; AAF67370.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX
 PS Claim 17: SEQ ID NO: 5905; 246bp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 XX Sequence 359 AA;
 SQ
 Query Match 25.2%; Score 56; DB 22; Length 359;
 Best Local Similarity 45.5%; Pred. No. 4.7;
 Matches 10; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 OY 1 PLEIWTIVPVILVIFGAFSL 22
 Db 108 pletivtlvplllvmvlfctv 129
 RESULT 13
 AAB79573
 ID AAB79573 standard; Protein; 359 AA.
 XX
 XX AAB79573;
 AC
 XX 30-APR-2001 (first entry)
 DT
 XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:662.
 DE
 XX Corynebacterium glutamicum; carbon metabolism and energy production;
 KW

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinoogenic amino acid;
 KW nonproteinoogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diptheriae; evolutionary study.
 XX
 XX Corynebacterium glutamicum.
 OS
 XX WO200100844-A2.
 PN
 XX 04-JAN-2001.
 PD
 XX 23-JUN-2000; 2000MO-IB00943.
 PF
 XX 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99US-0143208.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1040766.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX
 XX (BADI) BASF AG.
 PA
 XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 PI
 XX WPI: 2001-061975/07.
 DR N-PSDB; AAF71690.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 PT
 XX
 PS Claim 20; Page 1081-1082; 1246bp; English.
 PS
 XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinoenic
 CC or nonproteinoenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to



Db 61 METWTLPAIILVILALPSRLIWTDEVNPSLTIKSIG 101

RESULT 2

US-08-415-751-7

Sequence 7, Application US/08415751

Patent No. 5643772

GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEECH, JAMES

APPLICANT: NELSON, RICHARD, C.

TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-

TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA

TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID

TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND

TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESS: PHILLIPS, MOORE, LEMPIO & FINLEY

STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto

STATE: California

COUNTRY: United States of America

ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage

COMPUTER: PC

OPERATING SYSTEM: DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/415,751

FILING DATE: 03-APR-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/071,880

FILING DATE: June 1, 1993

APPLICATION NUMBER: 07/891,301

FILING DATE: May 29, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Hana Dolezalova

REGISTRATION NUMBER: 30,518

REFERENCE/DOCKET NUMBER: 480.19-2 (HMD)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-1677

TELEFAX: (415) 324-1678

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 361 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Cryptosporidium parvum

FEATURE:

NAME/KEY: Positions coded by nonsense codons are

NAME/KEY: Identified as Xaa.

US-08-415-751-7

Query Match 27.5%; Score 61; DB 1; Length 361;

Best Local Similarity 34.8%; Pred. No. 0.5;

Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

OY 3 ELVWTVIPVILVIFGAFSLPVL 25

Db 242 ELWMLLPXIVVVVGLGLPVL 264

RESULT 3

US-08-990-823-113

Sequence 113, Application US/08990823D

Patent No. 6228371

GENERAL INFORMATION:

APPLICANT: Nano, Francis

TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding

FILE REFERENCE: 49086

CURRENT FILING DATE: 1997-12-15

EARLIER FILING DATE: 1996-06-14

EARLIER FILING DATE: 1995-06-15

NUMBER OF SEQ ID NOS: 113

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 113

LENGTH: 363

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-08-990-823-113

Query Match 24.1%; Score 53.5; DB 4; Length 363;
Best Local Similarity 31.7%; Pred. No. 6.1;
Matches 13; Conservative 13; Mismatches 14; Indels 1; Gaps 1;OY 1 PLEIVTVIPVILVIFGAFSLPVLKQ-QEPPGDIYINV 40
Db 115 PLEIVTVIPVILVIFGAFSLPVLKQ-QEPPGDIYINV 155

RESULT 4

US-08-505-218-6

Sequence 6, Application US/08505218

Patent No. 5914447

GENERAL INFORMATION:

APPLICANT: ARAYA, ALEJANDRO

TITLE OF INVENTION: TRANSGENIC PLANTS INCLUDING A HYBRID

TITLE OF INVENTION: NUCLEIC ACID SEQUENCE, COMPRISING AN UNEDITED

TITLE OF INVENTION: MITOCHONDRIAL GENE FRAGMENT FROM HIGHER PLANTS AND PROCESS

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESS: LOWE, PRICE, LEBLANC & BECKER

STREET: 99 Canal Center Plaza, Suite 300

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/505,218

FILING DATE: 03-NOV-1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: KRAUS, ERIC J

REGISTRATION NUMBER: 36,190

REFERENCE/DOCKET NUMBER: 1169-013

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 684-1111

TELEFAX: (703) 684-1124

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 322 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-505-218-6

Query Match 23.9%; Score 53; DB 2; Length 322;
Best Local Similarity 48.0%; Pred. No. 6.2;
Matches 12; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 LEIVTVIVPVLVIFGAFSLPVLF 26
DB 145 MEIRITFVSIVLIFAIIPSFALLY 169

RESULT 5
US-07-759-568-1

Sequence 1, Application US/07759568
Patent No. 5374506
GENERAL INFORMATION:
APPLICANT: Murphy, Philip M.
TITLE OF INVENTION: Cloning of cDNA Encoding a Functional
TITLE OF INVENTION: Human Interleukin-8 Receptor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cushman, Darby & Cushman
STREET: 1615 L Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/759,568
FILING DATE: 19910913
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Scott, Watson T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: WTS/5683/91535/WBH
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 cush
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-759-568-1

Query Match 22.5%; Score 50; DB 1; Length 355;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVIVLIFGAFSLPVLF 27
DB 216 IVPVILMFCYGFLLRFLFK 235

RESULT 6
US-08-450-393A-8

Sequence 8, Application US/08450393A
Patent No. 5707815
GENERAL INFORMATION:
APPLICANT: Charo, Israel
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOTACTIC
TITLE OF INVENTION: PROTEIN RECEPTORS
NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306-2155

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,393A
FILING DATE: May 25, 1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: UCAL-237/0205
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5165
TELEFAX: 415-8857-0663
TELEX: 380816COOLEYPA
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-450-393A-8

Query Match 22.5%; Score 50; DB 1; Length 355;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVIVLIFGAFSLPVLF 27
DB 216 IVPVILMFCYGFLLRFLFK 235

RESULT 7
US-08-390-000A-5
Sequence 5, Application US/08390000A
Patent No. 5985583
GENERAL INFORMATION:
APPLICANT: Sealton, Stuart C.
TITLE OF INVENTION: Cloning and Expression of
TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penile & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,000A
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misticok, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-052
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9900
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-390-000A-5

Query Match 22.5%; Score 50; DB 2; Length 355;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIFGAFSLPVLFK 27
|||::|:| 1:1 |||
Db 216 IVPILIMFCYGFTRLTLFK 235

RESULT 8
US-08-446-669-8
Sequence 8, Application US/08446669
Patent No. 6132987
GENERAL INFORMATION:
APPLICANT: Chemo, Israel
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOATTRACTANT
NUMBER OF INVENTIONS: PROTEIN RECEPTORS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306-2155
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,669
FILING DATE: May 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Neeley, Richard
REGISTRATION NUMBER: 30,092
REFERENCE/DOCKET NUMBER: UCAL-237/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5000
TELEFAX: 415-857-0663
TELEX: 380816COOLEYPA
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-446-669-8

Query Match 22.5%; Score 50; DB 4; Length 355;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIFGAFSLPVLFK 27
|||::|:| 1:1 |||
Db 216 IVPILIMFCYGFTRLTLFK 235

RESULT 9
PCT-US95-00476-8
Sequence 8, Application PC/TUS9500476
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOATTRACTANT
NUMBER OF INVENTIONS: PROTEIN RECEPTORS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robbins, Berliner & Carson
STREET: 201 N. Figueroa Street, 5th Floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012-2628
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00476
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Berliner, Robert
REGISTRATION NUMBER: 20,121
REFERENCE/DOCKET NUMBER: 5555-291
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-977-1001
TELEFAX: 310-977-1003
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
PCT-US95-00476-8

Query Match 22.5%; Score 50; DB 5; Length 355;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 8 IVPVILVIFGAFSLPVLFK 27
|||::|:| 1:1 |||
Db 216 IVPILIMFCYGFTRLTLFK 235

RESULT 10
US-08-202-056-7
Sequence 7, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible

ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29, 135
REFERENCE/DOCKET NUMBER: 317743-108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 23:

QY 3 ELVWITVPVLLVETGAFSL---PVLFRQOEEFEGDVI 38
::| | | | | : : : | :
Db 75 KICWAEFTPTILITFLCSEFYQWEDMTGYSRYRPNMSWL 114

Thu Dec 27 08:21:48 2001

us-09-712-768-4.ra1

Page 7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:02:25 ; Search time 24.27 Seconds
(Without alignments)
138,100 Million cell updates/sec

Title: US-09-712-768-4
Perfect score: 222
Sequence: 1 PLEIWTIVPVILVIFGAF.....LFKQEPFGDIVINVEGRS 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	159	71.6	302	2 J01013	cytochrome-c oxida
2	139.5	62.8	297	1 OBPC2N	cytochrome-c oxida
3	96	43.2	313	2 E71698	cytochrome-c oxida
4	90	40.5	253	2 T12405	cytochrome-c oxida
5	89	40.1	229	2 S50328	cytochrome-c oxida
6	87	39.2	227	2 T11183	cytochrome-c oxida
7	86	38.7	227	2 S10190	cytochrome-c oxida
8	86	38.7	227	2 T11327	cytochrome-c oxida
9	86	38.7	227	2 T11025	cytochrome-c oxida
10	85	38.3	227	2 T11078	cytochrome-c oxida
11	85	38.3	229	2 T11170	cytochrome-c oxida
12	85	38.3	229	2 T12412	cytochrome-c oxida
13	84	37.8	230	2 S36009	cytochrome-c oxida
14	83	37.4	229	2 T11796	cytochrome-c oxida
15	83	37.4	230	2 T11537	cytochrome-c oxida
16	83	37.4	230	2 T11290	cytochrome-c oxida
17	82	36.9	227	2 T11196	cytochrome-c oxida
18	82	36.9	230	2 S68131	cytochrome-c oxida
19	82	36.9	230	2 T11167	cytochrome-c oxida
20	82	36.9	230	2 T09860	cytochrome-c oxida
21	82	36.9	230	2 T09950	cytochrome-c oxida
22	82	36.9	230	2 T11303	cytochrome-c oxida
23	81	36.5	225	2 T11155	cytochrome-c oxida
24	81	36.5	225	2 S70599	cytochrome-c oxida
25	81	36.5	230	2 S45491	cytochrome-c oxida
26	81	36.5	232	2 T11893	cytochrome-c oxida
27	81	36.5	228	2 T11893	cytochrome-c oxida
28	79	35.6	228	2 C38941	cytochrome-c oxida
29	79	35.6	228	2 T11130	cytochrome-c oxida

30	79	35.6	229	2 A45170	cytochrome-c oxida
31	79	35.6	231	2 I84422	cytochrome-c oxida
32	79	35.6	234	2 S35465	cytochrome-c oxida
33	79	35.6	237	2 S26949	cytochrome-c oxida
34	78	35.1	224	2 F45170	cytochrome-c oxida
35	78	35.1	228	1 OBEP2	cytochrome-c oxida
36	78	35.1	228	1 OBEP2Y	cytochrome-c oxida
37	78	35.1	228	1 T12000	cytochrome-c oxida
38	78	35.1	231	2 I36906	cytochrome-c oxida
39	77	34.7	226	2 T11352	cytochrome-c oxida
40	77	34.7	227	2 I62728	cytochrome-c oxida
41	77	34.7	230	2 T11457	cytochrome-c oxida
42	77	34.7	258	2 T11923	cytochrome-c oxida
43	77	34.7	873	1 S53828	cytochrome-c oxida
44	76	34.2	228	2 B38941	cytochrome-c oxida
45	76	34.2	228	2 T09802	cytochrome-c oxida

ALIGNMENTS

RESULT 1

J01013
cytochrome-c oxidase (EC 1.9.3.1) chain II precursor - Rhodobacter sphaeroides
C:Species: Rhodobacter sphaeroides
C:Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 23-Jul-1999
C:Accession: J01013
R:Gao, J.; Shapleigh, J.; Gennis, R.; Revzin, A.; Ferguson-Miller, S.
Gene 101, 133-137, 1991
A:Title: The gene encoding cytochrome c oxidase subunit II from Rhodobacter sphaeroid
A:Reference number: J01013; MUID:91285423
A:Accession: J01013
A:Molecule type: DNA
A:Residues: 1-302 <CAO>
A:Cross-references: GB:M57680; NID:g151893; PIDN:AAA26100.1; PID:g151894
C:Genetics:
A:Gene: coxII
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; oxidoreduct
F:1-24/Domain: (or 1-25) signal sequence #status predicted <SIG>
F:25-302/Product: (or 26-302) cytochrome-c oxidase polypeptide II #status predicted <
F:43-269/Domain: cytochrome-c oxidase chain II homology <CO2>
F:217,251,255,262/Binding site: copper 1 (His, Cys, Met) #status predicted
F:251,253,255,259/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:253/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 71.6%; Score 159; DB 2; Length 302;
Best Local Similarity 73.8%; Pred. No. 1,4e-12;
Matches 31; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PLEIWTIVPVILVIFGAFSLVLFKQEPFGDIVINVEG 42
Db 99 PLEIWTIVPVILVIFGAFSLVLFKQEPFGDIVINVEG 140

RESULT 2

OBPC2N
cytochrome-c oxidase (EC 1.9.3.1) chain II precursor - Paracoccus denitrificans
C:Species: Paracoccus denitrificans
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Jul-1999
C:Accession: S00106; S20231; S03803
R:Steinmeyer, P.; Steffens, G.C.M.; Panskus, G.; Buse, G.; Ludwig, B.
Eur. J. Biochem. 167, 431-439, 1987
A:Title: Subunit II of cytochrome c oxidase from Paracoccus denitrificans. DNA sequen
A:Reference number: S00106; MUID:88004464
A:Accession: S00106
A:Molecule type: DNA
A:Residues: 1-297 <STE>
A:Cross-references: EMBL:X05934; NID:g45483; PIDN:CAA29372.1; PID:g45484
A:Accession: S20231
A:Molecule type: protein
A:Residues: 29; B:261-280 <STE2>

R:Raitio, M.; Jalli, T.; Saraste, M.
EMBL J. 6, 2825-2833, 1987
A:Title: Isolation and analysis of the genes for cytochrome c oxidase in *Paracoccus denitrificans*
A:Reference number: S03803
A:Accession: S03803
A:Molecule type: DNA
A:Residues: 1-158, 'GV', 161-297 <RA1>
A:Cross-references: EMBL:X05828; NID:g45468; PIDN:CAA29268.1; PID:g45469
C:Genetics:
A:Gene: COII
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; oxidoreductase; pyro
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-280/Product: cytochrome-c oxidase chain II #status experimental <MAT>
F:29-54/Domain: periplasmic #status predicted <PER1>
F:48-662/Domain: cytochrome-c oxidase chain II homology <CO2>
F:55-87/Domain: transmembrane #status experimental <TM1>
F:102-133/Domain: transmembrane #status experimental <TM2>
F:134-280/Domain: carboxyl-terminal propylidene #status predicted <CTP>
F:281-297/Domain: carboxyl-terminal propylidene #status predicted <CTP>
F:209-244,248,255/Binding site: pyrrolidone carboxylic acid (Glu) (in mature form) #status experim
F:244,246,248,252/Binding site: copper 1 (His, Cys, Cys, Met) #status experimental
F:246/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 62.8%; Score 139.5; DB 1; Length 297;
Best Local Similarity 60.5%; Pred. No. 3,8e-10;
Matches 26; Conservative 9; Mismatches 7; Indels 1; Gaps 1;
OY 1 LEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 42
DB 104 PREIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 146

RESULT 3
E71698
cytochrome-c oxidase (EC 1.9.3.1) chain II RP406 - *Rickettsia prowazekii*
C:Species: *Rickettsia prowazekii*
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: E71698
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: E71698
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-313 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14863.1; PID:g386096
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: COXB; RP406
C:Superfamily: cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; oxidoreductase; resp
F:76-286/Domain: cytochrome-c oxidase chain II homology <CO2>
F:333,268,272,279/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:268,270,272,276/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:270/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 43.2%; Score 96; DB 2; Length 313;
Best Local Similarity 41.5%; Pred. No. 0.00012;
Matches 17; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
OY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 42
DB 133 IEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 173

RESULT 4
T12405
cytochrome-c oxidase (EC 1.9.3.1) chain II - *Sarcophyton glaucum* mitochondrion

C:Species: mitochondrion *Sarcophyton glaucum*
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jan-2000
C:Accession: T12405
R:Beaton, M.J.; Roger, A.J.; Cavaller-Smith, T.
J. Mol. Evol. 47, 697-708, 1998
A:Title: Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*: Conser
A:Reference number: Z17505; MUID:99065763
A:Accession: T12405
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-253 <BEA>
A:Cross-references: EMBL:AF064823; NID:g4091912; PID:g4091916; PIDN:AAC99650.1
C:Genetics:
A:Genome: mitochondrion
A:Note: COII
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrion
F:25-237/Domain: cytochrome-c oxidase chain II homology <CO2>
F:184,219,223,230/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:219,221,223,227/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:221/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 40.5%; Score 90; DB 2; Length 253;
Best Local Similarity 43.9%; Pred. No. 0.00053;
Matches 18; Conservative 7; Mismatches 16; Indels 0; Gaps 0;
OY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 42
DB 77 IEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 117

RESULT 5
S50328
cytochrome-c oxidase (EC 1.9.3.1) chain II - *Katharina tunnicata* mitochondrion
C:Species: mitochondrion *Katharina tunnicata*
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-Dec-1999
C:Accession: S50328
R:Boore, J.L.; Brown, W.M.
Genetics 138, 423-443, 1994
A:Title: Complete DNA sequence of the mitochondrial genome of the black chiton, *Katha*
A:Reference number: S50327; MUID:95129806
A:Accession: S50328
A:Molecule type: DNA
A:Residues: 1-229 <BOO>
A:Cross-references: EMBL:U09810; NID:g557273; PIDN:AAC48365.1; PID:g557277
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC4
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrion
F:9-214/Domain: cytochrome-c oxidase chain II homology <CO2>
F:161,196,200,207/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:196,198,200,204/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:198/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 40.1%; Score 89; DB 2; Length 229;
Best Local Similarity 43.9%; Pred. No. 0.00065;
Matches 18; Conservative 8; Mismatches 15; Indels 0; Gaps 0;
OY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 42
DB 61 IEIWTIVPVILVIFGAFSLPVLFKQEPPEGDIIVNEG 101

RESULT 6
T11183
cytochrome-c oxidase (EC 1.9.3.1) chain II - *Smithornis sharpei* mitochondrion
C:Species: mitochondrion *Smithornis sharpei*
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T11183

R:Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: Z17242
A:Accession: T11183
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-227 <MIND>
A:Cross-references: EMBL:AF090340; NID:g489448; PID:g4894492; PIDN:AA032517.1
C:Genetics:
A:Genome: mitochondrion
A:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 39.7%; Score 86; DB 2; Length 227;
Best Local Similarity 36.6%; Pred. No. 0.0011;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFTGAFSLPVLEFKQEPFGDIVINVEG 42
Db 60 VELWITLPAVLVLMALPSQILYMDIEDPDLTKAIG 100

RESULT 7
cytochrome-c oxidase (EC 1.9.3.1) chain II - chicken mitochondrion
C:Species: mitochondrion Gallus gallus (chicken)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 07-Dec-1999
C:Accession: S10190
R:Desjardins, P.; Morais, R.
J. Mol. Biol. 212, 599-634, 1990
A:Title: Sequence and gene organization of the chicken mitochondrial genome. A novel get
A:Reference number: S10187; MUID:90230301
A:Accession: S10190
A:Molecule type: DNA
A:Residues: 1-227 <DESS>
A:Cross-references: EMBL:X52392; NID:g12960; PIDN:CAA36628.1; PID:g12964
C:Genetics:
A:Gene: COII
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 38.7%; Score 86; DB 2; Length 227;
Best Local Similarity 34.1%; Pred. No. 0.0015;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFTGAFSLPVLEFKQEPFGDIVINVEG 42
Db 60 VELWITLPAVLVLMALPSQILYMDIEDPDLTKAIG 100

RESULT 8
cytochrome-c oxidase (EC 1.9.3.1) chain II - Corvus frugilegus mitochondrion
C:Species: mitochondrion Corvus frugilegus
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T11327
R:Harlid, A.; Arnason, U.
Proc. R. Soc. Lond. B Biol. Sci. 266, 305-309, 1999

A:Title: Analyses of mitochondrial DNA nest ratite birds within the Neognathae-support
A:Reference number: Z17262
A:Accession: T11327
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-227 <HAR>
A:Cross-references: EMBL:Y18522; PIDN:CAA77198.1
C:Genetics:
A:Genome: mitochondrion
A:Note: COII
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inn
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 38.7%; Score 86; DB 2; Length 227;
Best Local Similarity 36.6%; Pred. No. 0.0015;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFTGAFSLPVLEFKQEPFGDIVINVEG 42
Db 60 VELWITLPAVLVLMALPSQILYMDIEDPDLTKAIG 100

RESULT 9
cytochrome-c oxidase (EC 1.9.3.1) chain II - redhead mitochondrion
C:Species: mitochondrion Aythya americana (redhead)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T11025
R:Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: Z17242
A:Accession: T11025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-228 <JOH>
A:Cross-references: EMBL:AF090337; NID:g4887659; PID:g4887663; PIDN:AA032255.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inn
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 38.7%; Score 86; DB 2; Length 228;
Best Local Similarity 34.1%; Pred. No. 0.0015;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFTGAFSLPVLEFKQEPFGDIVINVEG 42
Db 60 VELWITLPAVLVLMALPSQILYMDIEDPDLTKAIG 100

RESULT 10
cytochrome-c oxidase (EC 1.9.3.1) chain II - Falco peregrinus mitochondrion
C:Species: mitochondrion Falco peregrinus
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C:Accession: T11078
R:Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: Z17242
A:Accession: T11078

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-227 <MIN>
A:Cross-references: EMBL:AF090338; NID:g4894462; PID:g4894466; PIDN:AAD32495.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match          38.3%; Score 85; DB 2; Length 227;
Best Local Similarity 31.7%; Pred. No. 0.002;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFICAFSLPVLFKQEPPEGDIVINVEG 42
DB 60 VELIWTILPAIVLILALPSLQILYIMDEIDPDLTAKIG 100

RESULT 11
T11170
Cytochrome-c oxidase (EC 1.9.3.1) chain II - greater rhea mitochondrion
C:Species: mitochondrion Rhea americana (greater rhea, common rhea)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T11170; T11419
R:Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 96:10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: Z17242
A:Accession: T11170
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <MIN>
A:Cross-references: EMBL:AF090339; NID:g4894475; PID:g4894478; PIDN:AAD32505.1
R:Harlid, A.
Submitted to the EMBL Data Library, May 1999
A:Reference number: Z17270
A:Accession: T11419
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <HAR>
A:Cross-references: EMBL:Y16884; PIDN:CAA76504.1
C:Genetics:
A:Gene: COII
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial
ein
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match          38.3%; Score 85; DB 2; Length 229;
Best Local Similarity 31.7%; Pred. No. 0.002;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFICAFSLPVLFKQEPPEGDIVINVEG 42
DB 60 VELIWTILPAIVLILALPSLQILYIMDEIDPDLTAKIG 100

RESULT 12
T12412
Cytochrome-c oxidase (EC 1.9.3.1) chain II - ostrich mitochondrion
C:Species: mitochondrion Struthio camelus (ostrich)

```

```

C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C:Accession: T12412; T11522
R:Mindell, D.P.; Sorenson, M.S.; Huddleston, C.J.; Miranda Jr., H.C.; Knight, A.; Saw
submitted to the EMBL Data Library, June 1998
A:Description: Phylogenetic relationships among and within select avian orders based
A:Reference number: Z17507
A:Accession: T12412
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <MIN>
A:Cross-references: EMBL:AF069429; NID:g3283618; PID:g3283625; PIDN:AAD09386.1
R:Harlid, A.
Submitted to the EMBL Data Library, September 1998
A:Reference number: Z17278
A:Accession: T11522
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-229 <HAR>
A:Cross-references: EMBL:Y12025; PIDN:CAA72747.1
C:Genetics:
A:Gene: COII
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inn
F:9-213/Domain: cytochrome-c oxidase chain II homology <CO2>
F:160,195,199,206/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:195,197,199,203/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:197/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match          38.3%; Score 85; DB 2; Length 229;
Best Local Similarity 31.7%; Pred. No. 0.002;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

QY 2 LEIWTIVPVILVFICAFSLPVLFKQEPPEGDIVINVEG 42
DB 60 VELIWTILPAIVLILALPSLQILYIMDEIDPDLTAKIG 100

RESULT 13
S36009
Cytochrome-c oxidase (EC 1.9.3.1) chain II - common carp mitochondrion
C:Species: mitochondrion Cyprinus carpio (common carp)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 23-Mar-2001
C:Accession: S36009; D44650
R:Chang, Y.S.; Huang, F.L.
Submitted to the EMBL Data Library, July 1991
A:Description: The cDNA and primary structure of pregrowth hormones of three species
A:Reference number: S21910
A:Accession: S36009
A:Molecule type: DNA
A:Residues: 1-230 <CHAI>
A:Cross-references: EMBL:X61010; NID:g436882; PIDN:CAA43340.1; PID:g12842
R:Chang, Y.S.; Huang, F.L.; Lo, T.B.
J. Mol. Evol. 38, 138-155, 1994
A:Title: The complete nucleotide sequence and gene organization of carp (Cyprinus car
A:Reference number: A44650; MUID:94222691
A:Accession: D44650
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-219, 'F', 221-230 <CHA2>
A:Cross-references: EMBL:X61010; NID:g436882
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial
ein
F:9-214/Domain: cytochrome-c oxidase chain II homology <CO2>
F:161,196,200,207/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted
F:196,198,200,204/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted
F:198/Binding site: magnesium (Glu) (shared with chain I) #status predicted

```

[illegible]

UDD CLINE. 427 DEC

Query Match	37.4%	Score	83	DB	2	Length	229
Best Local Similarity	41.5%	Pred. No.	0.0037				
Matches	17	Conservative	7	Mismatches	17	Indels	0
				Gaps	0		

QY	2	LEI	W	T	V	V	P	A	I	L	F	G	A	S	L	P	L	F	K	O	E	P	E	D	I	N	E	G	42
Db	61	LEI	T	V	V	P	A	I	L	F	G	A	S	L	P	L	F	K	O	E	P	E	D	I	N	E	G	101	

RESULT 15

T11537

cytochrome-c oxidase (EC 1.9.3.1) chain II - spiny dogfish mitochondrion (fragment)

C:Species: mitochondrion Squalus acanthias (spiny dogfish)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T11537

R:Rasmussen, A.S.; Arnason, U.

J. Mol. Evol. 48, 118-123, 1999

A:Title: Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous

A:Reference number: Z17281; MUID:99091711

A:Accession: T11537

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-230 <RMS>

A:Cross-references: EMBL:Y18134; NID:g4186095; PIDN:CAA77052.1; PID:g4186099

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SSCI

C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology

C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner

F:9-214/Domain: cytochrome-c oxidase chain II homology <CO2>

F:161,196,200,207/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted

F:196,199,200,204/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted

F:198/Binding site: magnesium (Glu) (shared with chain I) #status predicted

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:08:07 : Search time 15.26 Seconds
(without alignments)
105.718 Million cell updates/sec

Title: US-09-712-768-4

Perfect score: 222
Sequence: 1 PLEIWTIVPVVILVFIGNA.....LFKQEPPEGDIVINEGRS 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	159	71.6	302	1	COX2_RHOSH
2	139.5	62.8	298	1	COX2_PARDE
3	96	43.2	313	1	COX2_RICPR
4	86	38.7	199	1	COX2_APTAU
5	86	38.7	227	1	COX2_CHICK
6	86	38.7	227	1	COX2_COTLA
7	86	38.7	228	1	COX2_ANAPL
8	86	38.7	228	1	COX2_CAIMO
9	85	38.3	198	1	COX2_NOTPE
10	85	38.3	198	1	COX2_TINMA
11	85	38.3	199	1	COX2_CASBE
12	85	38.3	199	1	COX2_DRONO
13	85	38.3	199	1	COX2_RHEAM
14	85	38.3	229	1	COX2_SYRCA
15	84	37.8	230	1	COX2_CARAU
16	84	37.8	230	1	COX2_CYPCA
17	83	37.4	230	1	COX2_SQUAC
18	82	36.9	230	1	COX2_ONCMY
19	82	36.9	230	1	COX2_SALSA
20	82	36.9	230	1	COX2_SCYCA
21	81	36.5	225	1	COX2_RHISA
22	81	36.5	229	1	COX2_ASTPE
23	81	36.5	229	1	COX2_MXGL
24	81	36.5	230	1	COX2_GADMO
25	81	36.5	248	1	COX2_MESE
26	80	36.0	228	1	COX2_ABDAB
27	79	35.6	216	1	COX2_CAGO
28	79	35.6	228	1	COX2_ZOON
29	79	35.6	229	1	COX2_DROAI
30	79	35.6	229	1	COX2_DROAM
31	79	35.6	229	1	COX2_DROBF
32	79	35.6	229	1	COX2_DROLO
33	79	35.6	229	1	COX2_DRONR

34	79	35.6	229	1	COX2_DROPS	P29864 drosophila
35	79	35.6	229	1	COX2_DROSU	P29865 drosophila
36	79	35.6	230	1	COX2_CROLA	P34189 crossostoma
37	79	35.6	231	1	COX2_LAGLA	P98036 lagothrix 1
38	79	35.6	237	1	COX2_TRIRO	O01536 trichophyto
39	79	35.6	224	1	COX2_EXERO	P29873 exeristes r
40	78	35.1	228	1	COX2_ANOOU	P33505 anophelies q
41	78	35.1	228	1	COX2_DROME	P00408 drosophila
42	78	35.1	228	1	COX2_DROST	P50253 drosophila
43	78	35.1	229	1	COX2_DROYA	P00409 drosophila
44	78	35.1	228	1	COX2_STMTI	P98021 stimulum vi
45	78	35.1	231	1	COX2_ALOPA	P98024 alouatta pa

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	302 AA.
COX2_RHOSH				
ID	003736:			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CYTOCHROME C OXIDASE POLYPEPTIDE II PRECURSOR (EC 1.9.3.1)			
DE	(CYTOCHROME A3 SUBUNIT 2) (OXIDASE AA(3) SUBUNIT 2).			
GN	CTAC OR CTAB OR COXII.			
OS	Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;			
OC	Rhodobacter.			
OX	NCBI_TaxID=1063;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
KA	MEDLINE-91285423: PubMed=1648008:			
KA	Cao J., Shapleigh J., Gennis R., Revzin A., Ferguson-Miller S.,			
RT	"The gene encoding cytochrome c oxidase subunit II from Rhodobacter			
RT	sphaeroides; comparison of the deduced amino acid sequence with			
RT	sequences of corresponding peptides from other species."			
RL	Gene 101:133-137(1991).			
CC	-1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME			
CC	COMPLEX. ELECTRON TRANSFER IN CYTOCHROME C ARE TRANSFERRED VIA			
CC	HEME A AND CU(II) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND			
CC	CU(II).			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +			
CC	4 FERRICYTOCHROME C.			
CC	-1- COPROCTOR: COPPER A.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	EMBL: M57680; AAA26100.1; -			
DR	PIR: J01013; J01013.			
DR	HSSP: P08306; IAR1.			
DR	InterPro: IPR001505; COX2.			
DR	InterPro: IPR002429; Cyt_c_ox_2.			
DR	Pfam: PF00116; COX2; 1.			
DR	ProDom: PD000131; COX2; 1.			
DR	PROSITE: PS00078; COX2; 1.			
KW	Oxidoreductase; Respiratory chain; Electron transport; Transmembrane;			
KW	Copper; Signal.			
KW	Copper: Signal.			
FT	SIGNAL	1	24	POTENTIAL.
FT	CHAIN	25	302	CYTOCHROME C OXIDASE POLYPEPTIDE II.
FT	TRANSMEM	60	80	POTENTIAL.
FT	TRANSMEM	104	124	POTENTIAL.
FT	METAL	217	217	COPPER A (PROBABLE).
FT	METAL	251	251	COPPER A (PROBABLE).

FT METAL 255 255 COPPER A (PROBABLE).
 FT METAL 259 259 COPPER A (PROBABLE).
 SQ SEQUENCE 302 AA; 32940 MW; 4D40751A8E4C9AA4 CRC64;

Query Match 71.6%; Score 159; DB 1; Length 302;
 Best Local Similarity 73.8%; Pred. No. 1,6e-12;
 Matches 31; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 PLEIWTIVPVVILVIGAFSLPVLFGKQDFEPGDIVINVEG 42
 DB 99 PLEIAWTIVPVILVIGAFSLPVLFGKQDFEPGDIVINVEG 140

RESULT 2
 COX2_PARDE STANDARD; PRT; 298 AA.

ID COX2_PARDE STANDARD; PRT; 298 AA.
 AC P08306;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II PRECURSOR (EC 1.9.3.1)
 DE (CYTOCHROME A3 SUBUNIT 2) (OXIDASE AA(3) SUBUNIT 2).
 GN CTAC OR CTAB OR COIL.
 OS Paracoccus denitrificans.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Paracoccus.
 OX NCBI_TaxID=266;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S1657;
 RA Raitio M., Jalli T., Saraste M.;
 RT "Isolation and analysis of the genes for cytochrome c oxidase in
 RT Paracoccus denitrificans.";
 RL EMBL J. 6:2825-2833(1987).
 RN [2]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 13543;
 RA MEDLINE=88004464; PubMed=2820725;
 RA Steinuecke P., Steffens G.C.M., Penskus G., Buse G., Ludwig B.;
 RT "Subunit II of cytochrome c oxidase from Paracoccus denitrificans.
 RT DNA sequence, gene expression and the protein.";
 RL Eur. J. Biochem. 167:431-439(1987).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RA MEDLINE=95379947; PubMed=7651515;
 RA Iwata S., Ostermeier C., Ludwig B., Michel H.;
 RT "Structure at 2.8-A resolution of cytochrome c oxidase from
 RT Paracoccus denitrificans.";
 RL Nature 376:660-669(1995).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RA MEDLINE=98021406; PubMed=9380672;
 RA Ostermeier C., Harrenga A., Ernler U., Michel H.;
 RT "Structure at 2.7-A resolution of the Paracoccus denitrificans two-
 RT subunit cytochrome c oxidase complexed with an antibody FV
 RT fragment.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:10547-10553(1997).
 CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME
 CC COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA
 CC HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND
 CC CU(B).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERRICYTOCHROME C.
 CC -1- COFACTOR: TWO COPPER A ATOMS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X05828; CAA29268.1; ..
 DR PIR: S00106; OBP2N.
 DR PIR: IAR1; 11-FEB-96.
 DR InterPro: IPR001505; COX2.
 DR InterPro: IPR002429; cyL_c-ox_2.
 DR Pfam: PF00116; COX2; 1.
 DR PRINTS: PR01166; CYCOXIDASEII.
 DR ProDom: PD000131; COX2; 1.
 DR PROSITE: PS00078; COX2; 1.
 KW Oxidoreductase; Respiratory chain; Electron transport; Transmembrane;
 KW Copper; Signal; Inner membrane; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 280 CYTOCHROME C OXIDASE POLYPEPTIDE II.
 FT PROPEP 281 298 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 30 55 PERIPLASMIC.
 FT TRANSMEM 56 88 CYTOPLASMIC.
 FT DOMAIN 89 103
 FT TRANSMEM 104 134
 FT DOMAIN 135 280
 FT MOD_RES 30 30 PERIPLASMIC.
 FT METAL 210 210 PYROLDONE CARBOXYLIC ACID.
 FT METAL 245 245 COPPER A 1.
 FT METAL 247 247 COPPER A 1 AND 2.
 FT METAL 249 249 COPPER A 2.
 FT METAL 253 253 COPPER A 1 AND 2.
 FT METAL 256 256 COPPER A 1.
 FT CONFLICT 160 161 GV -> AF (IN REF. 2).
 SQ SEQUENCE 298 AA; 32539 MW; 918A64A9E3B93366 CRC64;

Query Match 62.8%; Score 139.5; DB 1; Length 298;
 Best Local Similarity 60.5%; Pred. No. 4e-10;
 Matches 26; Conservative 9; Mismatches 7; Indels 1; Gaps 1;

QY 1 PLEIWTIVPVVILVIGAFSLPVLFGKQDFEPGDIVINVEG 42
 DB 105 PLEIWTIVPVVILVIGAFSLPVLFGKQDFEPGDIVINVEG 147

RESULT 3
 COX2_RICPR STANDARD; PRT; 313 AA.
 ID COX2_RICPR STANDARD; PRT; 313 AA.
 AC Q92DC6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (CYTOCHROME
 DE A3 SUBUNIT 2).
 GN CTAC OR COXB OR RP406.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E;
 RA MEDLINE=99039499; PubMed=9823893;
 RA Anderson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Sticheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN [2]
 RP DOMAIN RPE1.
 RA MEDLINE=20485642; PubMed=11030655;
 RA Ogata H., Audic S., Barbe V., Artiguenave F., Fournier P.-E.,
 RA Raoult D., Claverie J.-M.;
 RT "Selfish DNA in protein-coding genes of Rickettsia.";
 RL Science 290:347-350(2000).

```

CC CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME
CC CC COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA
CC CC HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND
CC CC CU(B) (BY SIMILARITY).
CC CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC CC FERROCYTOCHROME C.
CC CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).
CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC CC -1- SIMILARITY: CONTAINS 1 REPEL INSERT DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: AJ235271; CAA1463.1; -.
CC CC InterPro: IPR001505; COX2.
CC CC InterPro: IPR002429; Cyl_c-ox_2.
CC CC Pfam: PF00116; COX2; 1.
CC CC DR PRINTS: PR01166; CYCOXIDASEII.
CC CC DR PRODOM: PD000131; COX2; 1.
CC CC DR PROSITE: PS00078; COX2; 1.
CC CC KW Oxidoreductase; Respiratory chain; Electron transport; Transmembrane;
CC CC Heme; Copper; Complete proteome.
CC CC FT TRANSMEM 39 59 POTENTIAL.
CC CC FT TRANSMEM 94 114 POTENTIAL.
CC CC FT TRANSMEM 131 151 POTENTIAL.
CC CC FT DOMAIN 5 51 REPEL INSERT.
CC CC FT METAL 233 233 COPPER A (POTENTIAL).
CC CC FT METAL 268 268 COPPER A (POTENTIAL).
CC CC FT METAL 272 272 COPPER A (POTENTIAL).
CC CC FT METAL 276 276 COPPER A (POTENTIAL).
CC CC SQ SEQUENCE 313 AA; 36037 MW; 77CDAD8F6A4201E5 CRC64;

Query Match 43.2%; Score 96; DB 1; Length 313;
Best Local Similarity 41.5%; Pred. No. 9.5e-05;
Matches 17; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

OY 2 LEIYTVIPVYVILVFAGSLPVLFKQOEPFGDIVINVG 42
Db 133 LEIYTVIPVILVILAVPSFRILRAEKIPEDLTIKYVG 173

RESULT 4
COX2_APTAU STANDARD; PRT: 199 AA.
ID COX2_APTAU
AC 003889;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MTCO2 OR COII.
OS Apterix australis (Brown kwl).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Apterygiformes; Apterygidae;
OC Apteryx.
OX NCBI_TaxID=8822;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";
RL (in) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-

```

```

CC CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.
CC CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC CC 4 FERROCYTOCHROME C.
CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC CC INNER MEMBRANE.
CC CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: U76064; AAB61314.1; -.
CC CC InterPro: IPR001505; COX2.
CC CC InterPro: IPR002429; Cyl_c-ox_2.
CC CC Pfam: PF00116; COX2; 1.
CC CC DR PRODOM: PD000131; COX2; 1.
CC CC DR PROSITE: PS00078; COX2; 1.
CC CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC CC Electron transport; Respiratory chain.
CC CC FT TRANSMEM 1 15 POTENTIAL.
CC CC FT DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
CC CC FT TRANSMEM 30 49 POTENTIAL.
CC CC FT TRANSMEM 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC CC FT METAL 128 128 COPPER A (PROBABLE).
CC CC FT METAL 163 163 COPPER A (PROBABLE).
CC CC FT METAL 167 167 COPPER A (PROBABLE).
CC CC FT METAL 171 171 COPPER A (PROBABLE).
CC CC SQ SEQUENCE 199 AA; 22374 MW; 0AF496B02F92ACB1 CRC64;

Query Match 38.7%; Score 86; DB 1; Length 199;
Best Local Similarity 34.1%; Pred. No. 0.001;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIYTVIPVYVILVFAGSLPVLFKQOEPFGDIVINVG 42
Db 28 VELVWTLIPALVILALPISLIYIMDEIDEDLTUKAIG 68

RESULT 5
COX2_CHICK STANDARD; PRT: 227 AA.
ID COX2_CHICK
AC P18944;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN MTCO2 OR COII.
OS Gallus gallus (chicken).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:90230301; PubMed:2329578;
RA Desjardins P., Morais R.;
RT "Sequence and gene organization of the chicken mitochondrial genome.
RT A novel gene order in higher vertebrates.";
RL J. Mol. Biol. 212:599-634(1990)
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.

```

```

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52392; CAA36628.1; -.
CC DR PIR: S10190; S10190.
CC DR HSSP: P00404; 10CC.
CC DR InterPro: IPR001505; COX2.
CC DR InterPro: IPR002429; Cyt_c-ox_2.
CC DR Pfam: PF00116; COX2; 1.
CC DR PRINTS: PR01166; CYCOXIDASEII.
CC DR ProDom: PD000131; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT TRANSMEM 27 47 POTENTIAL.
CC FT DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 62 81 POTENTIAL.
CC FT DOMAIN 82 227 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 160 160 COPPER A (PROBABLE).
CC FT METAL 195 195 COPPER A (PROBABLE).
CC FT METAL 199 199 COPPER A (PROBABLE).
CC FT METAL 203 203 COPPER A (PROBABLE).
CC SQ SEQUENCE 227 AA; 25568 MW; 7B0C1CEBCD72AED CRC64;

Query Match 38.7%; Score 86; DB 1; Length 227;
Best Local Similarity 34.1%; Pred. No. 0.0012;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

Qy 2 LEIWTIVPVILVIFGAFSLPVLFKQOEPFGDIYINVG 42
Db 60 VELIWTILPAIVLVLLALPSLIQLYIMDEIDEPDLTLKAIG 100

RESULT 6
COX2_COTJA STANDARD: PRT: 227 AA.
AC P50684;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN COII.
OS Coturnix coturnix japonica (Japanese quail).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCB1_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Ramirez V., Morais R.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.

```

```

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U36794; AA76730.1; -.
CC DR HSSP: P00404; 10CC.
CC DR InterPro: IPR001505; COX2.
CC DR InterPro: IPR002429; Cyt_c-ox_2.
CC DR Pfam: PF00116; COX2; 1.
CC DR PRINTS: PR01166; CYCOXIDASEII.
CC DR ProDom: PD000131; COX2; 1.
CC DR PROSITE: PS00078; COX2; 1.
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT TRANSMEM 27 47 POTENTIAL.
CC FT DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 62 81 POTENTIAL.
CC FT DOMAIN 82 227 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 160 160 COPPER A (PROBABLE).
CC FT METAL 195 195 COPPER A (PROBABLE).
CC FT METAL 199 199 COPPER A (PROBABLE).
CC FT METAL 203 203 COPPER A (PROBABLE).
CC SQ SEQUENCE 227 AA; 25607 MW; 3BF1E03A46455AFE CRC64;

Query Match 38.7%; Score 86; DB 1; Length 227;
Best Local Similarity 34.1%; Pred. No. 0.0012;
Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;

Qy 2 LEIWTIVPVILVIFGAFSLPVLFKQOEPFGDIYINVG 42
Db 60 VELIWTILPAIVLVLLALPSLIQLYIMDEIDEPDLTLKAIG 100

RESULT 7
COX2_ANAPL STANDARD: PRT: 228 AA.
AC P98019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN COII.
OS Anas platyrhynchos (Domestic duck).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCB1_TaxID=8839;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEKIN BRED; TISSUE=Liver;
RX MEDLINE=94047124; PubMed=8230253;
RA Ramirez V., Savoie P., Morais R.;
RT "Molecular characterization and evolution of a duck mitochondrial
RT genome.";
RL J. Mol. Evol. 37:296-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PEKIN BRED; TISSUE=Liver;
RA Pan Y.F., Lee Y.H.W., Wei Y.H., Chiang A.N.;
RT "A gene cytochrome c oxidase subunit II in duck mitochondrial DNA:
RT structural features and sequence evolution.";
RL Biochem. Int. 30:479-489(1993).
RN [3]

```

RP SEQUENCE OF 22-54 FROM N.A.
 RC STRAIN=PEKIN BREED; TISSUE=Liver;
 RX MEDLINE=90360426; PubMed=2390786;
 RA Desjardins P., Ramirez V., Morais R.;
 RT "Gene organization of the Pekin duck mitochondrial genome.";
 RL Curr. Genet. 17:515-518(1990).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERRICYTOCHROME C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L22476; AAA72039.1; -;
 DR EMBL: X68507; CAA48519.1; -;
 DR EMBL: X55531; CAA39147.1; -;
 DR HSSP: P00404; 10CC.
 DR InterPro: IPR001505; COX2.
 DR InterPro: IPR002429; Cyt_c-ox-2.
 DR Pfam: PF00116; COX2; 1.
 DR PRINTS: PR01166; CYCOXIDASE1.
 DR PRODOM: PD000131; COX2; 1.
 DR PROSITE: PS00078; COX2; 1.
 KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
 KW Electron transport; Respiratory chain.
 FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT TRANSSEM 27 47 POTENTIAL.
 FT DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
 FT TRANSSEM 62 81 POTENTIAL.
 FT DOMAIN 82 228 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT METAL 160 160 COPPER A (PROBABLE).
 FT METAL 195 195 COPPER A (PROBABLE).
 FT METAL 199 199 COPPER A (PROBABLE).
 FT METAL 203 203 COPPER A (PROBABLE).
 SQ SEQUENCE 228 AA; 25553 MW; E1C4FE82797DBE03 CRC64;

Query Match 38.7%; Score 86; DB 1; Length 228;
 Best Local Similarity 34.1%; Pred. No. 0.0012;
 Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;
 QY 2 LEIWTIYPIVILVFICAFSLPVLFKQGEFPEGDIVINVG 42
 DB 60 VELMTITLPAIVLVLLALPSLQIILYMDDEIDEDPLITKAIG 100

RESULT 8
 COX2_CAIMO STANDARD; PRT: 228 AA.
 AC P50666;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
 GN COI1.
 OS Cairina moschata (Muscovy duck).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Cairina.
 OX NCBI_TaxID=8855;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MUSCOVY SHINNEY DUCK; TISSUE=Liver;
 RA Pan Y.F., Lee Y.H.W., Wei Y.H., Chiang A.N.;
 RT "A gene cytochrome C oxidase subunit II in duck mitochondrial DNA:
 structural features and sequence evolution.";
 RL Biochem. Int. 30:479-489(1993).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERRICYTOCHROME C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X68508; CAA48520.1; -;
 DR HSSP: P00404; 10CC.
 DR InterPro: IPR001505; COX2.
 DR InterPro: IPR002429; Cyt_c-ox-2.
 DR Pfam: PF00116; COX2; 1.
 DR PRINTS: PR01166; CYCOXIDASE1.
 DR PRODOM: PD000131; COX2; 1.
 DR PROSITE: PS00078; COX2; 1.
 KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
 KW Electron transport; Respiratory chain.
 FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT TRANSSEM 27 47 POTENTIAL.
 FT DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
 FT TRANSSEM 62 81 POTENTIAL.
 FT DOMAIN 82 228 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 FT METAL 160 160 COPPER A (PROBABLE).
 FT METAL 195 195 COPPER A (PROBABLE).
 FT METAL 199 199 COPPER A (PROBABLE).
 FT METAL 203 203 COPPER A (PROBABLE).
 SQ SEQUENCE 228 AA; 25523 MW; D34E9E3797DBCBB CRC64;

Query Match 38.7%; Score 86; DB 1; Length 228;
 Best Local Similarity 34.1%; Pred. No. 0.0012;
 Matches 14; Conservative 11; Mismatches 16; Indels 0; Gaps 0;
 QY 2 LEIWTIYPIVILVFICAFSLPVLFKQGEFPEGDIVINVG 42
 DB 60 VELMTITLPAIVLVLLALPSLQIILYMDDEIDEDPLITKAIG 100

RESULT 9
 COX2_NOTPE STANDARD; PRT: 198 AA.
 AC 003892;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
 GN MTCO2 OR COI1.
 OS Notoprocta perdicaria (Chilean tinamou).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae;
 OC Notoprocta.
 OX NCBI_TaxID=30464;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RL between molecular and morphological data sets.";
RL (In) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76067; AAB61323.1; -.
CC DR InterPro: IPR001505; COX2.
CC DR InterPro: IPR002429; Cyt_c-ox_2.
CC DR Pfam: PF00116; COX2; 1.
CC DR ProDom: PD000131; COX2; 1.
CC DR ProSITE: PS00078; COX2; 1.
CC KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT NON_TER 1 1
CC FT TRANSMEM <1 15 POTENTIAL.
CC FT DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 30 49 POTENTIAL.
CC FT DOMAIN 50 198 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 128 128 COPPER A (PROBABLE).
CC FT METAL 163 163 COPPER A (PROBABLE).
CC FT METAL 167 167 COPPER A (PROBABLE).
CC FT METAL 171 171 COPPER A (PROBABLE).
CC SQ SEQUENCE 198 AA; 22227 MW; 7A27344A157E6A13 CRC64;

Query Match 38.3%; Score 85; DB 1; Length 198;
Best Local Similarity 31.7%; Pred. No. 0.0014;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIGAFSLPYLFKQDEPPEGDIVINVEG 42
DB 28 VELIWTILPAVLILALPLSLQILLYMDEIDEPDLTKAIG 68

RESULT 10
COX2_TINMA STANDARD; PRT; 198 AA.
ID COX2_TINMA STANDARD; PRT; 198 AA.
AC 003895;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MFCO2 OR COII.
OS Tinamus major (Great tinamou).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae; Tinamus.
OX NCBI_TaxID=30468;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";

```

```

RL (In) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76070; AAB61332.1; -.
CC DR HSPB, P08306; 1A81.
CC DR InterPro: IPR001505; COX2.
CC DR InterPro: IPR002429; Cyt_c-ox_2.
CC DR Pfam: PF00116; COX2; 1.
CC DR ProDom: PD000131; COX2; 1.
CC DR ProSITE: PS00078; COX2; 1.
CC KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT NON_TER 1 1
CC FT TRANSMEM <1 15 POTENTIAL.
CC FT DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 30 49 POTENTIAL.
CC FT DOMAIN 50 198 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 128 128 COPPER A (PROBABLE).
CC FT METAL 163 163 COPPER A (PROBABLE).
CC FT METAL 167 167 COPPER A (PROBABLE).
CC FT METAL 171 171 COPPER A (PROBABLE).
CC SQ SEQUENCE 198 AA; 22305 MW; E655564EED39BA89 CRC64;

Query Match 38.3%; Score 85; DB 1; Length 198;
Best Local Similarity 31.7%; Pred. No. 0.0014;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIGAFSLPYLFKQDEPPEGDIVINVEG 42
DB 28 VELIWTILPAVLILALPLSLQILLYMDEIDEPDLTKAIG 68

RESULT 11
COX2_CASBE STANDARD; PRT; 199 AA.
ID COX2_CASBE STANDARD; PRT; 199 AA.
AC 003890;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN MFCO2 OR COII.
OS Casuaris bennetti (Dwarf cassowary).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Casuariiformes; Casuariidae;
OC Casuaris.
OX NCBI_TaxID=30463;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K., Feinstein J., Cracraft J.;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RT between molecular and morphological data sets.";
RL (In) Mindell D.P. (eds.);

```

RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
 RL New York (1997).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERRICYTOCHROME C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U76065; AAB61317.1; .
 CC InterPro: IPR001505; COX2.
 CC Pfam: PF00116; COX2; 1; Cyt_c-ox_2.
 CC ProDom: PD000131; COX2; 1.
 CC PROSITE: PS00078; COX2; 1.
 CC Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
 CC Electron transport; Respiratory chain.
 CC NON_TER 1
 CC TRANSMEM <1 15 POTENTIAL.
 CC DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
 CC TRANSMEM 30 49 POTENTIAL.
 CC DOMAIN 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 CC METAL 128 128 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 CC METAL 163 163 COPPER A (PROBABLE).
 CC METAL 167 167 COPPER A (PROBABLE).
 CC METAL 171 171 COPPER A (PROBABLE).
 CC SEQUENCE 199 AA; 22378 MW; 62E1DAE5E57C92E8 CRC64;
 SQ
 Query Match 38.3%; Score 85; DB 1; Length 199;
 Best Local Similarity 31.7%; Pred. No. 0.0014;
 Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;
 Oy 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFEGDIVNVEG 42
 Db 28 VELIWTILPAIVLILALPSLQILYMDIEDPDLTKAIG 68
 RESULT 12
 COX2_DRONO STANDARD; PRT; 199 AA.
 AC 003891;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
 GN MTCO2 OR COII.
 OS Rhea americana (Greater rhea) (Common rhea).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Palaeognathae; Casuariiformes; Dromadidae;
 CC Dromasus
 CC NCBI_TaxID=8790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lee K., Feinstein J., Cracraft J.;
 RT "Phylogenetic relationships of the ratite birds: resolving conflicts
 RT between molecular and morphological data sets."
 RL (in) Mindell D.P. (eds.);
 RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
 RL New York (1997).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERRICYTOCHROME C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U76066; AAB61320.1; .
 CC InterPro: IPR001505; COX2.
 CC Pfam: PF00116; COX2; 1; Cyt_c-ox_2.
 CC ProDom: PD000131; COX2; 1.
 CC PROSITE: PS00078; COX2; 1.
 CC Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
 CC Electron transport; Respiratory chain.
 CC NON_TER 1
 CC TRANSMEM <1 15 POTENTIAL.
 CC DOMAIN 16 29 MITOCHONDRIAL MATRIX (POTENTIAL).
 CC TRANSMEM 30 49 POTENTIAL.
 CC DOMAIN 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 CC METAL 128 128 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
 CC METAL 163 163 COPPER A (PROBABLE).
 CC METAL 167 167 COPPER A (PROBABLE).
 CC METAL 171 171 COPPER A (PROBABLE).
 CC SEQUENCE 199 AA; 22336 MW; 62E1DAE557C92E8 CRC64;
 SQ
 Query Match 38.3%; Score 85; DB 1; Length 199;
 Best Local Similarity 31.7%; Pred. No. 0.0014;
 Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;
 Oy 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFEGDIVNVEG 42
 Db 28 VELIWTILPAIVLILALPSLQILYMDIEDPDLTKAIG 68
 RESULT 13
 COX2_RHEAM STANDARD; PRT; 199 AA.
 AC 003893;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
 GN MTCO2 OR COII.
 OS Rhea americana (Greater rhea) (Common rhea).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Palaeognathae; Rheiformes; Rheidae; Rhea.
 CC NCBI_TaxID=8797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lee K., Feinstein J., Cracraft J.;
 RT "Phylogenetic relationships of the ratite birds: resolving conflicts
 RT between molecular and morphological data sets."
 RL (in) Mindell D.P. (eds.);
 RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
 RL New York (1997).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2

CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -I- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; MITOCHONDRIAL
CC INNER MEMBRANE.
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.stb.ch).
CC -----
DR EMBL: U76068; AAB61326.1; -
DR InterPro: IPR001505; COX2.
DR DR InterPro: IPR002429; Cyt_c-cox_2.
DR Pfam: PF00116; COX2; 1.
DR ProDom: PD000131; COX2; 1.
DR ProSITE: PS00078; COX2; 1.
KM Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
KV Electron transport; Respiratory chain.
FT NON_TER 1
FT TRANSMEM 1
FT POTENTIAL.
FT DOMAIN <1 15 MITOCHONDRIAL MATRIX (POTENTIAL).
FT TRANSMEM 16 29 POTENTIAL.
FT TRANSMEM 30 49 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT METAL 50 199 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT METAL 128 128 COPPER A (PROBABLE).
FT METAL 163 163 COPPER A (PROBABLE).
FT METAL 167 167 COPPER A (PROBABLE).
FT METAL 171 171 COPPER A (PROBABLE).
SQ SEQUENCE 199 AA; 22265 MW; 58725E39529773B7 CRC64;

Query Match 38.3%; Score 85; DB 1; Length 199;
Best Local Similarity 31.7%; Pred. No. 0.0014;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

Oy 2 LEIWTIVPVVLVIFGAFSLPELFKQGFPEGDIVINVEG 42
 :::|||||::: 11:: 11:: 11:: 11:
Db 28 VELWTLPAIVLILLALPSIQIYLMMDEIDEDLTAKIG 68

RESULT 14
COX2_STRCA STANDARD; PRT; 229 AA.

AC 021400; 003894;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN MTCO2 OR COIT.
OS Struthio camelus (Ostrich).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Struthioniformes; Struthionidae;
OC Struthio.
OX NCBI_TaxID=8801;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97357422; PubMed=9214748;
RA Harlid A., Janke A., Arnason U.;
RT "The mtDNA sequence of the ostrich and the divergence between
RT palaeonathous and neognathous birds.";
RL Mol. Biol. Evol. 14:754-761(1997).
LN [2]
RP SEQUENCE FROM N.A.
RA Sorenson M.D., Dimcheff D.E., Ast J.C., Yuri T., Mindell D.P.;
RT "Primers for a PCR-based approach to complete mitochondrial genome
RT sequencing".
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.

```

RN  [3]
RP  SEQUENCE OF 33-229 FROM N.A.
RA  Lee K., Feinstein J., Cracraft J.:
RT  "Phylogenetic relationships of the ratite birds: resolving conflicts
RL  between molecular and morphological data sets.";
RL  (in) Mindell D.P. (eds.);
RL  Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL  New York (1997).
CC  -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC  CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC  3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC  TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINDING COPPER
CC  A CENTER TO THE BINETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC  -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC  4 FERRICCYTOCHROME C.
CC  -1- SUBCELLULAR LOCATION: INTERAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC  INNER MEMBRANE.
CC  -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC  -----
CC  CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: Y12025; CAA72747.1; -.
DR  EMBL: AF069429; AAD09366.1; -.
DR  EMBL: U76069; AAB61329.1; ALT_TERM.
DR  HSSP: P08306; IAR1
DR  InterPro: IPR001505; COX2.
DR  InterPro: IPR002429; CYL_c-ox_2.
DR  Pfam: PF00116; COX2; 1.
DR  PRINTS: PR01166; CYCOXIDASEII.
DR  ProDom: PD000131; COX2; 1.
DR  PROSITE: PS00078; COX2; 1.
KW  Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
KW  Electron transport; Respiratory chain.
FT  DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT  TRANSMEM 27 47 POTENTIAL.
FT  DOMAIN 48 61 MITOCHONDRIAL MATRIX (POTENTIAL).
FT  TRANSMEM 62 81 POTENTIAL.
FT  DOMAIN 82 229 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
FT  METAL 160 160 COPPER A (PROBABLE).
FT  METAL 195 195 COPPER A (PROBABLE).
FT  METAL 199 199 COPPER A (PROBABLE).
FT  METAL 203 203 COPPER A (PROBABLE).
SQ  SEQUENCE 229 AA; 25699 MW; B9377617BACB14EF CRC64;

Query Match 38.3%; Score 85; DB 1; Length 229;
Best Local Similarity 31.7%; Pred. No. 0.0016;
Matches 13; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

Qy 2 LEIWTIVPVVILVIFGAFSLPVLFKQDFEGDIVINVEG 42
:|::|||:|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 60 VELIWTIIPALVILILALPLSLQIILYKMADEIDEPDITLKATG 100

RESULT 15
ID COX2_CARAU STANDARD; PRT: 230 AA.
AC 076882;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
OS COXII OR COII.
OS Carassius auratus (Goldfish).
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

```



```

CC Cypriniformes: Cyprinidae; Cyprininae; Carassius.
OX NCBI_taxid=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A23; TISSUE-Oocyte;
RA Murakami M., Yamashita Y., Fujitani H.;
RT "the complete sequence of mitochondrial genome from a gynogenetic
RT triploid 'gibuna' (Carassius auratus langsdorfi).";
RL Zool. Sci. 15:335-337(1998).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AB06953; BAA31241.1; -.
CC DR InterPro; IPR001505; COX2.
CC DR InterPro; IPR002429; CYL_C_OX_2.
CC DR Pfam; PF00116; COX2; 1.
CC DR PRINTS; PR01166; CYCOXIDASE1.
CC DR PRODom; PD000131; COX2; 1.
CC DR PROSITE; PS00078; COX2; 1.
CC KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;
CC Electron transport; Respiratory chain.
CC FT DOMAIN 1 26 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT TRANSMEM 27 48 PROBABLE.
CC FT DOMAIN 49 62 MITOCHONDRIAL MATRIX (POTENTIAL).
CC FT TRANSMEM 63 82 PROBABLE.
CC FT DOMAIN 83 230 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).
CC FT METAL 161 161 COPPER A (PROBABLE).
CC FT METAL 196 196 COPPER A (PROBABLE).
CC FT METAL 200 200 COPPER A (PROBABLE).
CC FT METAL 204 204 COPPER A (PROBABLE).
CC SEQUENCE 230 AA; 26045 MW; 39A0B54FE31B2377 CRC64;

```

```

Query Match 37.8%; Score 84; DB 1; Length 230;
Best Local Similarity 43.9%; Pred. No. 0.0021;
Matches 18; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

```

```

QY 2 LEIWTIVPVILVIFGAFSLPVLFKQEFPEGDIVINVEG 42
   :|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|
Db 61 IEIWTILPAVILVILVIALPSLRITILYLMDEINDPHLTIKAMG 101

```

Search completed: December 22, 2001, 11:08:07
Job time: 336 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:07:40 ; Search time 37.53 Seconds
(without alignments)
171.489 Million cell updates/sec

Title: US-09-712-768-4

Perfect score: 222
Sequence: 1 PLEIWTIVPVVILVIFGA.....LFKQEPPEGDIVINWGSR 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP:invertebrate:*
6: SP:mammal:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139.5	62.8	297	2	059663 paracoccus
2	100	45.0	309	2	09A229
3	99	44.6	285	2	059630 nitrobacter
4	92	41.4	279	2	09RM99 bradyrhizob
5	90	40.5	228	8	09B6T1 eudomia el
6	89	40.1	253	8	099619 sarcophyton
7	89	40.1	208	8	09GAL1 cryptocorne
8	89	40.1	229	8	037534 kalharina l
9	88	39.6	247	8	09TBV8 acropora te
10	87	39.2	208	8	09GAL6
11	87	39.2	227	8	09XU17 smthornis
12	87	39.2	227	8	09G5U8 lessonia ru
13	87	39.2	227	8	09G5U7
14	87	39.2	229	8	037584 orchessella
15	86	39.2	230	8	09G849 chrysomya n
16	86	38.7	226	8	09G2J4 phytomyza v
17	86	38.7	226	8	09G1H9 phytomyza v
18	86	38.7	227	8	09Z234 corvus frug
19	86	38.7	227	8	09TBH9 chalcites l

20	86	38.7	227	8	09TBH8	09LbH8 cacomantis
21	86	38.7	227	8	09TBH0	09LbH0 corythaixi
22	86	38.7	227	8	09TBG9	09LbG9 opisthocomu
23	86	38.7	227	8	09M91	09M91 scytalopus
24	86	38.7	227	8	09M90	09M90 scelorchillu
25	86	38.7	227	8	09M09	09M09 scelorchillu
26	86	38.7	227	8	09M08	09M08 pteropochro
27	86	38.7	227	8	09M07	09M07 pteropochro
28	86	38.7	227	8	09M06	09M06 pteropochro
29	86	38.7	227	8	09M05	09M05 pteropochro
30	86	38.7	227	8	09G5U9	09G5U9 tyrannus me
31	86	38.7	227	8	09G5U6	09G5U6 agriornis m
32	86	38.7	227	8	09G5U5	09G5U5 muscisaxico
33	86	38.7	227	8	09G5U4	09G5U4 muscisaxico
34	86	38.7	227	8	09G2G4	09G2G4 muscisaxico
35	86	38.7	227	8	09G2G1	09G2G1 muscisaxico
36	86	38.7	227	8	09G2G0	09G2G0 muscisaxico
37	86	38.7	228	8	09XK24	09XK24 aythya amer
38	86	38.7	228	8	09B701	09B701 apteryx man
39	85	38.3	96	8	09B980	09B980 blastophaga
40	85	38.3	227	8	09XLO5	09XLO5 falco pereg
41	85	38.3	227	8	09TB14	09TB14 turnix vari
42	85	38.3	227	8	09TB13	09TB13 chaetura pe
43	85	38.3	227	8	09TB12	09TB12 bubo virgin
44	85	38.3	227	8	09TB11	09TB11 chordellus
45	85	38.3	227	8	09TBH7	09TBH7 piaya melan

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	297 AA
059663	059663			
AC	059663			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CYTCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) PRECURSOR.			
OS	Paracoccus denitrificans.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;			
OC	Paracoccus.			
OX	NCBI_TaxID=266;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=WILD TYPE;			
RC	MEDLINE=88004464; PubMed=2820725;			
RA	Steinbecke P., Steffens G.C.M., Panskus G., Buse G., Ludwig B.;			
RT	"Subunit II of cytochrome c oxidase from Paracoccus denitrificans. DNA			
RT	sequence, gene expression and the protein."			
RL	Eur. J. Biochem. 167:431-439(1987)			
CC	-1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME			
CC	COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA			
CC	HEME A AND CU(A) TO THE BINDING CENTER FORMED BY HEME A3 AND			
CC	CU(B). SUBUNIT II BINDS CU(A) AND CYTOCHROME C.			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4			
CC	FERRICYTOCHROME C.			
CC	-1- COFACTOR: COPPER A AND HEME GROUP.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	PRODOM: PD000131; COX2.1.			
CC	PROSITE: PS00078; COX2.1.			
DR	HSSP: P08306; IAR1.			
DR	InterPro: IPR001505; COX2.			
DR	InterPro: IPR002429; Cyt_c-ox_2.			
DR	Pfam: PF00116; COX2.1.			
DR	PRINTS: PR01166; CYCOXIDASEII.			
DR	PRODOM: PD000131; COX2.1.			
DR	PROSITE: PS00078; COX2.1.			
KW	Copper; Oxidoreductase; Signal; Transmembrane.			
FT	SIGNAL	28		POTENTIAL.
FT	CHAIN	29		POTENTIAL.
FT	METAL	211		COPPER A (PROBABLE).
FT	METAL	215		COPPER A (PROBABLE).

"Extended sequencing of a DNA fragment of B. japonicum adjacent to the
 RT cox operon."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 DR EMBL: AJ242592; CAB56818.1; -
 DR EMBL: U33883; AAF78816.1; -
 DR HSSP: P18400; ICYX.
 DR InterPro: IPR001505; COX2.
 DR InterPro: IPR002429; Cyt_c_ox_2.
 DR Pfam: PF00116; COX2; 1
 DR PRINTS: PR01166; CYCOXIDASEIT.
 DR PRODOM: PD000131; COX2; 1.
 DR PROSITE: PS00078; COX2; 1.
 KW Copper; Oxidoreductase; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 SQ SEQUENCE 279 AA; 30564 MW; 5FA79C394FA73C8 CRC64;

Query Match 41.4%; Score 92; DB 2; Length 279;
 Best Local Similarity 40.5%; Pred. No. 0.0023;
 Matches 17; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 43
 DB 99 IEVNTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 140

RESULT 5

09B6T1 PRELIMINARY; PRT; 228 AA.

AC 09B6T1.
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DE CYTOCHROME C OXIDASE SUBUNIT II.
 OS Eudromia elegans (Elegant crested-tinamou).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae; Eudromia.
 OX NCBI_TaxID=8805;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BREAST MUSCLE;
 RX MEDLINE=21085666; PubMed=11217857;
 RA Cooper A., Lalueza-Fox C., Anderson S., Rambaut A., Austin J.,
 RA Ward R.;
 RT "Complete mitochondrial genome sequences of two extinct moas clarify
 RT rattle evolution.";
 RL Nature 409:704-707(2001).
 DR EMBL: AY016016; AAK08590.1; -
 KW Mitochondrion.
 SQ SEQUENCE 228 AA; 25668 MW; 26C2ECCAS813F9E3 CRC64;

Query Match 40.5%; Score 90; DB 8; Length 228;
 Best Local Similarity 33.3%; Pred. No. 0.0032;
 Matches 14; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 43
 DB 60 VELIWTIPALVILFIFGASLPVLFKQEPPEGDIVINVEGR 101

RESULT 6

09G6I9 PRELIMINARY; PRT; 253 AA.

AC 09G6I9.
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).

GN COIT
 OS Sarcophyton glaucum.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Aleyonaria; Aleyonacea;
 OC Aleyoniidae; Sarcophyton.
 OX NCBI_TaxID=70919;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99065763; PubMed=9847412;
 RA Beaton M.J., Roger A.J., Cavalier-Smith T.;
 RT "Sequence analysis of the mitochondrial genome of Sarcophyton
 RT conserved gene order among octocorals.";
 RL J. Mol. Evol. 47:697-708(1998).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINDING COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
 CC FERROCYTOCHROME C.
 CC -1- COFACTOR: COPPER A AND HEME GROUP.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 DR EMBL: AF064823; AAC99650.1; -
 DR HSSP: P08306; IARL.
 DR InterPro: IPR001505; COX2.
 DR InterPro: IPR002429; Cyt_c_ox_2.
 DR Pfam: PF00116; COX2; 1.
 DR PRINTS: PR01166; CYCOXIDASEIT.
 DR PRODOM: PD000131; COX2; 1.
 DR PROSITE: PS00078; COX2; 1.
 KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
 KN Respiratory chain; Transmembrane.
 FT METAL 186 186 COPPER A (PROBABLE).
 FT METAL 190 190 COPPER A (PROBABLE).
 FT METAL 230 230 COPPER A (PROBABLE).
 SQ SEQUENCE 253 AA; 28391 MW; F8B37D4C08939D0F CRC64;

Query Match 40.5%; Score 90; DB 8; Length 253;
 Best Local Similarity 43.9%; Pred. No. 0.0036;
 Matches 18; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVFIFGASLPVLFKQEPPEGDIVINVEGR 42
 DB 77 IEIWTIPALVILVFIFGASLPVLFKQEPPEGDIVINVEGR 117

RESULT 7

09G6I1 PRELIMINARY; PRT; 208 AA.

AC 09G6I1.
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
 GN COIL.
 OS Cryptotermes secundus.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Isoptera;
 OC Kalotermitidae; Cryptotermitinae; Cryptotermites.
 OX NCBI_TaxID=105785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Thompson G.J., Lenz M., Miller L., Crozier R.H.;
 RT "Phylogenetic Analysis and Trait Evolution in Australian Lineages of
 RT Drywood Termites (Isoptera, Kalotermitidae)."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-

3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY SIMILARITY).

-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4 FERROCYTOCHROME C.

-1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).

-1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.

EMBL: AF189093; AAC17077.1; -

Interpro: IPR001505; COX2.

DR Interpro: IPR002429; Cyt_c-ox-2.

DR Pfam: PF00116; COX2; 1.

DR PRINTS: PR01166; CYCOXIDASEII.

DR PRODOM: PD000131; COX2; 1.

DR PROSITE: PS00078; COX2; 1.

KW Copper; Inner membrane; Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane.

FT NON_TER 1 1

FT SEQUENCE 208 AA; 23745 MW; E787EDAD0D3D91987 CRC64;

Query Match 40.1%; Score 89; DB 8; Length 208;
Best Local Similarity 41.5%; Pred. No. 0.0039;
Matches 17; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIVINEG 42
Db 42 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIVINEG 82

RESULT 8
Q37534 PRELIMINARY; PRT; 229 AA.

ID Q37534
AC Q37534;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
OS Katharina tunicata (black chiton).
OG Mitochondrion.
OC Eukaryota; Metazoa; Mollusca; Polyplacophora; Neoloricata;
OC Ischnochitonina; Mopaliidae; Katharina.
OX NCBI_Taxid=34587;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95129806; Pubmed=7828825;
RA Boore J.L.; Brown W.M.;
RT "Complete DNA sequence of the mitochondrial genome of the black
RT chiton, Katharina tunicata";
RL Genetics 136:423-443(1994).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
CC SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERROCYTOCHROME C.

CC -1- COFACTOR: COPPER A AND HEME GROUP.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).

CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.

DR EMBL: U09810; AAC48365.1; -

DR HSSP: P18400; ICYX.

DR Interpro: IPR001505; COX2.

DR Interpro: IPR002429; Cyt_c-ox-2.

DR Pfam: PF00116; COX2; 1.

DR PRINTS: PR01166; CYCOXIDASEII.

DR PRODOM: PD000131; COX2; 1.

DR PROSITE: PS00078; COX2; 1.

KW Copper; Inner membrane; Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane.

FT SEQUENCE 247 AA; 28024 MW; 9F7D810C4B2F8D06 CRC64;

KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.

FT METAL 163 163 COPPER A (PROBABLE).
FT METAL 167 167 COPPER A (PROBABLE).
FT METAL 196 196 COPPER A (PROBABLE).
FT METAL 200 200 COPPER A (PROBABLE).
FT METAL 204 204 COPPER A (PROBABLE).
FT METAL 207 207 COPPER A (PROBABLE).
FT SEQUENCE 229 AA; 26079 MW; FA0BAB11965E754E CRC64;

Query Match 40.1%; Score 89; DB 8; Length 229;
Best Local Similarity 43.9%; Pred. No. 0.0043;
Matches 18; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIVINEG 42
Db 61 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIVINEG 101

RESULT 9
Q9TBV8 PRELIMINARY; PRT; 247 AA.

ID Q9TBV8
AC Q9TBV8;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
GN COIL.
OS Acropora tenuis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astroceniina; Acroporidae; Acropora.
OX NCBI_Taxid=70783;

RN [1]
RP SEQUENCE FROM N.A.
RA van Oppen M.J.H.; Hislop N.R.; Miller D.J.;
RT "The mitochondrial genome of the scleractinian coral Acropora tenuis:
RT major differences in gene order within the anthozoa subclass
RT Zoantharia";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
CC SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERROCYTOCHROME C.

CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).

CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.

DR EMBL: AF152244; AAD52920.1; -

DR HSSP: P18400; ICYX.

DR Interpro: IPR001505; COX2.

DR Interpro: IPR002429; Cyt_c-ox-2.

DR Pfam: PF00116; COX2; 1.

DR PRINTS: PR01166; CYCOXIDASEII.

DR PRODOM: PD000131; COX2; 1.

DR PROSITE: PS00078; COX2; 1.

KW Copper; Inner membrane; Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane.

FT SEQUENCE 247 AA; 28024 MW; 9F7D810C4B2F8D06 CRC64;

Query Match 39.6%; Score 88; DB 8; Length 247;
Best Local Similarity 46.3%; Pred. No. 0.0061;
Matches 19; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

OY 2 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIVINEG 42
Db 77 LEIWTIVPVILVIFGAFSLPVLFKQEPFGDIVINEG 117

```

RESULT 10
09GAL6 PRELIMINARY; PRT; 208 AA.
AC 09GAL6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
GN COIL.
OS Cryptococcus dudleyi.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Isoptera;
OC Kalotermitidae; Cryptotermitinae; Cryptotermeae.
OX NCBI_TaxID=105782;
RN [1]
RP SEQUENCE FROM N.A.
RA Thompson G.J., Lenz M., Miller L., Crozier R.H.;
RT "Phylogenetic Analysis and Trait Evolution in Australian Lineages of
RT Drywood Termites (Isoptera, Kalotermitidae).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL: AF189088; AAG17072.1; -.
DR InterPro: IPR001505; COX2.
DR InterPro: IPR002429; Cyl_c_ox_2.
DR Pfam: PF00116; COX2; 1.
DR PRINTS: PR01166; CYCOXIDASEII.
DR PRODOM: PD000131; COX2; 1.
DR PROSITE: PS00078; COX2; 1.
DR Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1
FT NON_TER 208
FT SEQUENCE 208 AA; 23739 MW; 3C87E913464B9813 CRC64;
SQ
Query Match 39.2%; Score 87; DB 8; Length 208;
Best Local Similarity 39.0%; Pred. No. 0.0068;
Matches 16; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
QY 2 LEIVWTIVPVILVIGAFSLPVLFKQEPPEGDIYINVEG 42
Db 42 IEVWMTIPALVILFTAMPSLRLYLMDLTHPALTLYKVG 82
.
RESULT 11
09XLI7 PRELIMINARY; PRT; 227 AA.
AC 09XLI7;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
OS Smithornis sharpei.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Eurylimidae;
OC Smithornis.
OX NCBI_TaxID=81930;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=98393707; Pubmed=9724766;
RA Mindell D.P., Sorenson M.D., Dimcheff D.E.;
RT "Multiple Independent Origins of Mitochondrial Gene Order in Birds.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10693-10697(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Mindell D.P., Sorenson M.D., Dimcheff D.E., Hasegawa M., Ast J.C.,
RT "Interordinal Relationships of Birds and Other Reptiles based on whole
RT mitochondrial genomes.";
RL Syst. Biol. 48:138-152(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Sorenson M.D., Dimcheff D.E., Ast J.C., Yuri T., Mindell D.P.;
RT "Complete mitochondrial DNA sequences for five birds and a turtle.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: COPPER A AND HEME GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL: AF090340; AAD32517.1; -.
DR InterPro: IPR001505; COX2.
DR InterPro: IPR002429; Cyl_c_ox_2.
DR Pfam: PF00116; COX2; 1.
DR PRINTS: PR01166; CYCOXIDASEII.
DR PRODOM: PD000131; COX2; 1.
DR PROSITE: PS00078; COX2; 1.
DR Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT METAL 162
FT METAL 166
FT METAL 206
FT METAL 206
FT SEQUENCE 227 AA; 25592 MW; 1CC3AC8ACE83AF96 CRC64;
SQ
Query Match 39.2%; Score 87; DB 8; Length 227;
Best Local Similarity 36.6%; Pred. No. 0.0074;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;
QY 2 LEIVWTIVPVILVIGAFSLPVLFKQEPPEGDIYINVEG 42
Db 60 VELVWTLIPALVILVMTALPSLQILYMDLTHPALTLYKAIG 100
.
RESULT 12
09G5U8 PRELIMINARY; PRT; 227 AA.
AC 09G5U8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1).
OS Lessonia rufa.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Tyrannidae; Lessonia.
OX NCBI_TaxID=123643;
RN [1]
RP SEQUENCE FROM N.A.
RA Chessier R.T.;
RT "Evolution in the high Andes: the phylogenetics of Muscisaxicola
RT ground-tyrants.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY

```


AC Q9G849;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
 GN COII.
 OS Chrysomya norrisi.
 OG Mitochondrion.
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Oestridae; Calliphoridae; Chrysomya.
 OX NCBI_TaxID=142904;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wells J.D., Sperling F.A.;
 RT "DNA Based Identification of Forensically Important Chrysomyinae
 (Diptera: Calliphoridae)".
 RL Forensic Sci. Int. 0:0-0(2001).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1 -
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
 CC FERRICYTOCHROME C.
 CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
 DR EMBL: AF295552; AAG34093.1; -
 DR InterPro: IPR001505; COX2.
 DR InterPro: IPR002429; CYL_g_ox_2.
 DR Pfam: PF00116; COX2; 1.
 DR PRINTS: PR01166; CYCOXIDASEII.
 DR ProDom: PD000131; COX2; 1.
 DR PROSITE: PS00078; COX2; 1.
 KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
 KW Respiratory chain; Transmembrane.
 FT NON_TER 230
 SQ SEQUENCE 230 AA; 26368 MW; 969373C75291F85C CRC64;

Query Match 39.2%; Score 87; DB 8; Length 230;
 Best Local Similarity 43.9%; Pred. No. 0.0075;
 Matches 18; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

OY 2 LEIWTIVPVVILVIGAFSLPYLFKQDEFPBGDIVINVEG 42
 :||:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
 DB 61 IEIWTILPAIILFLTAPPSLRLLYLDELINERSTLKVIG 101

Search completed: December 22, 2001, 11:07:40
 Job time: 349 sec

OM nucleic - nucleic search, using sw model
December 22, 2001, 09:44:42 : Search time 2710.12 Seconds
Run on: (without alignments)
803.517 Million cell updates/sec

Title: US-09-712-768-3
 Perfect score: 132
 Sequence: 1 ccgctggaatcgtctcgac.....tcaacgtcgaaggtctgaac 132

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Parameters: 2944280

Total number of hits satisfying chosen parameters

Minimum DB seq length:	0
Maximum DB seq length:	2000000000

post-processing:	Minimum	Match	0%
	Maximum	Match	100%

Database :

```

1:  GenEmb1: *
2:  1:  gb.ba: *
3:  2:  gb.htg: *
4:  3:  gb.in: *
5:  4:  gb.om: *
6:  5:  gb.ov: *
7:  6:  gb.par: *
8:  7:  gb.ph: *
9:  8:  gb.pl: *
10: 9:  gb.pr: *
11: 10: gb.ro: *
12: 11: gb.sts: *
13: 12: gb.sy: *
14: 13: gb.un: *
15: 14: gb.vi: *
16: 15: em.ba: *
17: 16: em.fun: *
18: 17: em.hum: *
19: 18: em.in: *
20: 19: em.om: *
21: 20: em.ov: *
22: 21: em.pa: *
23: 22: em.pl: *
24: 23: em.pl: *
25: 24: em.ro: *
26: 25: em.sts: *
27: 26: em.sy: *
28: 27: em.vi: *
29: 28: em.vi: *
30: 29: em.htg_hum: *
31: 30: em.htg_inv: *
32: 31: em.htg_rod: *
33: 32: em.htg_hum: *
34: 33: em.htg_inv: *
35: 34: em.htg_rod: *
36: 35: em.htg_other: *

```

30: $ch_{i,j} = \frac{1}{n} \sum_{k=1}^n \text{pred}_{i,j,k}$
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed.
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	132	100.0	132	6	AX113988	AX113988 Sequence
2	75.4	57.1	1356	1	M57680	M57680 Rhodobacter
3	48.2	36.5	1653	1	PDX0X1	X05934 P. denitrific
4	48.2	36.5	1653	1	PDX0X1	X05928 Paracoccus
5	47.6	36.1	8121	1	BJA242592	AJ242592.Bradyrhiz
6	47.6	35.1	31495	1	BJU33883	X33883 Bradyrhizob
7	47.2	32.8	11188	1	AEO06001	AEO06001 Caulobact
8	39.4	28.8	4380	1	NMCOX4AC	X89566 N. winograd
9	37.6	28.5	12829	1	AEO04449	AEO04449 Pseudom
10	36.2	27.4	5206	3	DMU36762	X56762 Drosophil
11	36.2	27.4	4500	3	DMU32073	AJ002073 Drosophi
12	36.2	27.4	4500	3	DMU32073	X89246 D.melanog
13	35.6	27.0	1260	3	SYNBTCTCONS	L074951 Bacterial
14	35.4	26.8	16537	5	AB047553	AB047553 Plectogl
15	34.2	25.9	311	9	HS7A8R	B61594 H.sapiens
16	33.6	25.5	316	9	HS197D1R	C257801 H.sapiens
17	33.6	25.5	684	9	HSU12698	U12690 Human Hsa2
18	33.6	25.5	16569	9	AF346988	AF346988 Homo sapi
19	33.2	25.2	624	3	AF189104	AF189104 Kalcitene
20	33.3	25.0	590	2	AC074564	AC074564 Gliardia
21	32.6	24.7	317	9	HS13F31R	C259274 H.sapiens
22	32.2	24.7	197219	2	AC087839	AC087839 Oryza sat
23	32.2	24.4	9810	4	AF018073	AF018073 Rhodobact
24	32.2	24.2	1886	6	MTMRGEN	U10524 Macropus ro
25	32.2	24.2	182	6	AX063146	AX063146 Sequence
26	32	24.2	195	6	AX192821	AX192821 Sequence
27	32	24.2	130	6	AX062955	AX062955 Sequence
28	32	24.2	236	6	AX192820	AX192820 Sequence
29	32	24.2	270	6	AX062842	AX062842 Sequence
30	32	24.2	272	6	AX063008	AX063008 Sequence
31	32	24.2	275	6	AX063006	AX063006 Sequence
32	32	24.2	275	6	AX063116	AX063116 Sequence
33	32	24.2	277	6	AX062826	AX062826 Sequence
34	32	24.2	281	6	AX063087	AX063087 Sequence
35	32	24.2	288	6	AX062978	AX062978 Sequence
36	32	24.2	289	6	AX062964	AX062964 Sequence
37	32	24.2	294	6	AX062954	AX062954 Sequence
38	32	24.2	306	6	AX062854	AX062854 Sequence
39	32	24.2	313	6	AB0192819	AX192819 Sequence
40	32	24.2	313	6	AB019563	AB019563 Homo sapi
41	32	24.2	586	9	MFHSCOX11	X356541 H.sapiens
42	32	24.2	586	9	AK026536	AK026536 Homo sapi
43	32	24.2	595	9	AF004339	AF004339 Homo sapi
44	32	24.2	684	9	HSU12691	U12691 Human Hsa3
45	32	24.2	684	9	HSU12691	U12691 Human Hsa3

RESULT	1		PAT	01-MAY-2001
AXI13988				
LOCUS	AXI13988	132 bp	DNA	
DEFINITION	Sequence 3 from Patent EP1103603.			
ACCESSION	AXI13988			
VERSION	AXI13988.1	GI:13940148		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS	1 (bases 1 to 132)			
TITLE	Asakura, A., Hoshino, T. and Shinjoh, M.			
JOURNAL	Cytochrome c oxidase complex from gluconobacter oxydans			
	Patent: EP 1103603-A 3 30-MAY-2001;			
	F. HOFMANN-LA ROCHE AG (CH)			
FEATURES				
source	1..132			
	location="gluconobacter oxydans"			

```
1. .132
/organism="Gluconobacter oxydans"
/db_xref="taxon:442"
```

Query Match	100.0%;	Score 132;	DB 6;	Length 132;
Best Local Similarity	100.0%;	Pred. No.	4.3e-26;	
Matches 132;	Conservative	0.		

RESULT 2

M57680.3 GI:15022166
 cytochrome c oxidase subunit II; transmembrane protein.
 Rhodobacter sphaeroides.
 Rhodobacter sphaeroides.
 ORGANISM
 SOURCE
 KEYWORDS
 CLASSIFICATION

Encodings of cytochrome c oxidase subunit II from *Rhodobacter rubrum* and *Rhodospirillum rubrum* (Miller, S. Sphaeroides): comparison of the deduced amino acid sequence with sequences of corresponding peptides from other species

JOURNAL
Gene 101 (1), 133-137 (1991)

MEDLINE
91285423

PUBMED
 1648008
 REFERENCE
 2 (bases 1 to 1356)
 AUTHORS
 Hiser, C. and Ferguson-Miller, S.
 TITLE
 Direct Submission
 JOURNAL

REMARK
COMMENT
FEATURES

submitted (26-JUL-2001) Biochemistry Department, Michigan State University, East Lansing, MI 48824, USA
Sequence update by submitter
On Jul 26, 2001 this sequence version replaced

```
source      Location/Qualifiers
1. .1356    replaced g1:15022158
```

```
gene      /db_xref="taxon:1063" spinaeroides
349..1260
CDS       /gene="coxII"
349..1260
```

```

/ gene="cox"
/ EC_number="1.9.3.1"
/ codon_start=1
/ trans1_table=1
/ product="cytochrome c oxidase subunit II"
/ protein_id="AAA26100.3"
/ db_xref="GI:15022167"
/ translation="MHSSTTTCGCATGAAGLAATAAAAGGSSLEIIIGPGCGIGF
PSASPVATQIHMIDGFIITATITTEIVLLIVYVMEFHRKRNQAPARTINSLEI
AMTIVPILVALAGAFSLPTEVNGDELPTLVVTVGMYKMYGVD"

```

Query Match	57.1%	Score	75.4;	DB	1;	Length	1356;
Best Local	75.2%	Pred.	No. 1.3e-10;				
Matches	94;	Conservative					

RESULT 3

	X05934	P. denitrificans oxid gene for cytochrome c oxidase subunit II (EC 1.9.3.1).	DNA	BCT	21-MAR-1995
DEFINITION	POOXI2	1653 bp			
LOCUS	H08003				
ACCESSION					
FEATURES					

X05934.1 GI:45483
 cytochrome c oxidase subunit II.
 Paracoccus denitrificans.
 Paracoccus denitrificans.
 ORGANISM
 SOURCE
 KEYWORDS
 UNCLON

REFERENCE AUTHORS TITLE	Paracoccus, I (bases 1 to 1653) Ludwig B. Direct Submission Submitted (31-AUG-1987, 11:11)	alpha subdivision; Rhodobacter group
-------------------------------	--	--------------------------------------

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES

Luebeck I Universitäts- und Landesbibliothek zu Köln, Universitätsstr. 160, 50907 Köln.
2 (bases 1 to 1653)
Steinrucke P., Steffens G.C., Panisk G., Buse G. and Ludwig B.
Subunit II of cytochrome c oxidase from Paracoccus denitrificans.
DNA sequence, gene expression and the protein
Eur J. Biochem. 167 (3), 431-439 (1987).
88004464

source	Location/Qualifiers
	1. .1653

```

DSDS      /strain="wild type"
           /db_xref="taxon:266"
           /clone_1fb="pUC8"
           /clone_470="
533. .615
           /product="signal peptide (AA -28 to 11"
sig_peptide

```

```

002. .1425      (11 20 10 -1)".
/note="precursor polypeptide (11 20 10 -1)".

```

```

/codon_start=1
/transl_table=1
/protein_id="CA29372.1"
/db_xref="GI:45484"
/db_xref="SPPREMBL:O59663"
/translation="MALAIRRGVY
GMNQPSSPLAHDQDLDHPVLYIAVNSLGVATRPVPAADVDGLDLPVYKGRK
TPIVEVITLPEVLLIVATGASPLPLLRSDMPNDPLVLTALGHQWNSYVYRTH
AFIDALMERKALDADGSESDLEPLTNDPPVAVPGKKLVQVATVYTHATITPAFA
KODAPVIRIQTALDAGSDRGVITLNDPPVAVPGKKLVQVATVYTHATITPAFA
EFADSDSYPPSPYKLSAE
616..1422

```

BASE COUNT		/product="mature cytochrome oxidase subunit II (AA 1-269)
270	a	
525	c	
556	g	
302	t	

gene	CDS	BASE COUNT	ORIGIN
gene	CDS	1282 a 2639 c 2619 g 1581 t	<pre> /transl_table=11 /evidence=experimental /product="putative heme A synthase or assembly protein" /protein_id="CAB56822.1" /db_xref="GI:6006416" /translat="MDHEPTIARDSDRTARKGVKGLGRDVLVASICGVALLMVGAS YAAPFVMEKCRATGENTGTQVATAAPGTGLAKIVREDSNADPLPMKFEPESE LEVINIGQVTFVTVINOAAPRTAGCAAVNAPLTVGSYFQKINCEFTFOTAPAEK RMDVAVYNDPSIADHDENDGLSTITLSVTFYPAKDPVYKPLASGEGRKKNL" 5981..6877 /gene="coxh" 5981..6877 /EC_number="1.9.3.1" /function="cytochrome c oxidation (respiration)" /codon_start=1 /transl_table=11 /product="cytochrome c oxidase subunit III" /protein_id="CAB56823.1" /db_xref="GI:6006417" /translat="MTATAGCKNHIDYILVDPSPWPAVGSLSAFINAVGALMMHMEFA APVAYGATFGVLTVMASWMDGVYIEQAKYGDHTFVVDLHNRYSMLITLSEVVFVA WPAVFNAAALFPRADAVNATPRADAVNATPRADAVNATPRADAVNATPRADAVN TLLTLLTGTTVVAHHALLLENDROGKRYKLIITVVLGALFTVCQAVAEVSHAASFEAGN LYGATFEPMATGFGFHVLTGTTFLVLCFLRAVAGHFTKQHLGFEFAWYIYHFPDYW LFLFLCYVWGHGCAETMAHGAH" 6975..7376 /EC_number="1.9.3.1" /function="cytochrome c oxidase subunit III" /protein_id="CAB56824.1" /db_xref="GI:6006418" /translat="MNDTGAPEPKSETPTVLQSMARGLACPCRGQGLYAGFLTLAASS CDRGDVAFLDIDGDPALFIMLAGIIVACALIVYKQPPYLLAHLVWLPLIPLAT TLPLPRAMKSLILALQPHKAAAGRLVDRAK" 7199..8121 /EC_number="1.9.3.1" /function="cytochrome a3 biogenesis" /codon_start=1 /transl_table=11 /product="SUR1 homolog" /protein_id="CAB56825.1" /db_xref="GI:6006419" /translat="MRADRGQPPAAVLAACRAVVAADBRHNAALATARDEVAADRAAI PROGAGPAGRPRENNETARKPRVAGFALTLLTAFAALGWQJORTAKHEILIAE LTERIAAPIALPPAPMAALNPRADEFREVSFTAPASPDMAVYSSGSAVRKAS POTNAFLPARLPSEGVYIDAGFVEMTMDORSDEDAVKKIYGVQVALTGILRFPEE PQMLTAPASRDKRLMFVFRVDYALIASLIGKGTVAAPYITDLEQAPRAGITRCPRLDYH KDDHLYAVTWFAAGAVLLAEVGVWGRSS" </pre>
gene	CDS	1282 a 2639 c 2619 g 1581 t	<pre> /transl_table=11 /evidence=experimental /product="putative heme A synthase or assembly protein" /protein_id="CAB56822.1" /db_xref="GI:6006416" /translat="MDHEPTIARDSDRTARKGVKGLGRDVLVASICGVALLMVGAS YAAPFVMEKCRATGENTGTQVATAAPGTGLAKIVREDSNADPLPMKFEPESE LEVINIGQVTFVTVINOAAPRTAGCAAVNAPLTVGSYFQKINCEFTFOTAPAEK RMDVAVYNDPSIADHDENDGLSTITLSVTFYPAKDPVYKPLASGEGRKKNL" 5981..6877 /EC_number="1.9.3.1" /function="cytochrome c oxidation (respiration)" /codon_start=1 /transl_table=11 /product="cytochrome c oxidase subunit III" /protein_id="CAB56823.1" /db_xref="GI:6006417" /translat="MTATAGCKNHIDYILVDPSPWPAVGSLSAFINAVGALMMHMEFA APVAYGATFGVLTVMASWMDGVYIEQAKYGDHTFVVDLHNRYSMLITLSEVVFVA WPAVFNAAALFPRADAVNATPRADAVNATPRADAVNATPRADAVNATPRADAVN TLLTLLTGTTVVAHHALLLENDROGKRYKLIITVVLGALFTVCQAVAEVSHAASFEAGN LYGATFEPMATGFGFHVLTGTTFLVLCFLRAVAGHFTKQHLGFEFAWYIYHFPDYW LFLFLCYVWGHGCAETMAHGAH" 6975..7376 /EC_number="1.9.3.1" /function="cytochrome c oxidase subunit III" /protein_id="CAB56824.1" /db_xref="GI:6006418" /translat="MNDTGAPEPKSETPTVLQSMARGLACPCRGQGLYAGFLTLAASS CDRGDVAFLDIDGDPALFIMLAGIIVACALIVYKQPPYLLAHLVWLPLIPLAT TLPLPRAMKSLILALQPHKAAAGRLVDRAK" 7199..8121 /EC_number="1.9.3.1" /function="cytochrome a3 biogenesis" /codon_start=1 /transl_table=11 /product="SUR1 homolog" /protein_id="CAB56825.1" /db_xref="GI:6006419" /translat="MRADRGQPPAAVLAACRAVVAADBRHNAALATARDEVAADRAAI PROGAGPAGRPRENNETARKPRVAGFALTLLTAFAALGWQJORTAKHEILIAE LTERIAAPIALPPAPMAALNPRADEFREVSFTAPASPDMAVYSSGSAVRKAS POTNAFLPARLPSEGVYIDAGFVEMTMDORSDEDAVKKIYGVQVALTGILRFPEE PQMLTAPASRDKRLMFVFRVDYALIASLIGKGTVAAPYITDLEQAPRAGITRCPRLDYH KDDHLYAVTWFAAGAVLLAEVGVWGRSS" </pre>
gene	CDS	1282 a 2639 c 2619 g 1581 t	<pre> /transl_table=11 /evidence=experimental /product="putative heme A synthase or assembly protein" /protein_id="CAB56822.1" /db_xref="GI:6006416" /translat="MDHEPTIARDSDRTARKGVKGLGRDVLVASICGVALLMVGAS YAAPFVMEKCRATGENTGTQVATAAPGTGLAKIVREDSNADPLPMKFEPESE LEVINIGQVTFVTVINOAAPRTAGCAAVNAPLTVGSYFQKINCEFTFOTAPAEK RMDVAVYNDPSIADHDENDGLSTITLSVTFYPAKDPVYKPLASGEGRKKNL" 5981..6877 /EC_number="1.9.3.1" /function="cytochrome c oxidation (respiration)" /codon_start=1 /transl_table=11 /product="cytochrome c oxidase subunit III" /protein_id="CAB56823.1" /db_xref="GI:6006417" /translat="MTATAGCKNHIDYILVDPSPWPAVGSLSAFINAVGALMMHMEFA APVAYGATFGVLTVMASWMDGVYIEQAKYGDHTFVVDLHNRYSMLITLSEVVFVA WPAVFNAAALFPRADAVNATPRADAVNATPRADAVNATPRADAVNATPRADAVN TLLTLLTGTTVVAHHALLLENDROGKRYKLIITVVLGALFTVCQAVAEVSHAASFEAGN LYGATFEPMATGFGFHVLTGTTFLVLCFLRAVAGHFTKQHLGFEFAWYIYHFPDYW LFLFLCYVWGHGCAETMAHGAH" 6975..7376 /EC_number="1.9.3.1" /function="cytochrome c oxidase subunit III" /protein_id="CAB56824.1" /db_xref="GI:6006418" /translat="MNDTGAPEPKSETPTVLQSMARGLACPCRGQGLYAGFLTLAASS CDRGDVAFLDIDGDPALFIMLAGIIVACALIVYKQPPYLLAHLVWLPLIPLAT TLPLPRAMKSLILALQPHKAAAGRLVDRAK" 7199..8121 /EC_number="1.9.3.1" /function="cytochrome a3 biogenesis" /codon_start=1 /transl_table=11 /product="SUR1 homolog" /protein_id="CAB56825.1" /db_xref="GI:6006419" /translat="MRADRGQPPAAVLAACRAVVAADBRHNAALATARDEVAADRAAI PROGAGPAGRPRENNETARKPRVAGFALTLLTAFAALGWQJORTAKHEILIAE LTERIAAPIALPPAPMAALNPRADEFREVSFTAPASPDMAVYSSGSAVRKAS POTNAFLPARLPSEGVYIDAGFVEMTMDORSDEDAVKKIYGVQVALTGILRFPEE PQMLTAPASRDKRLMFVFRVDYALIASLIGKGTVAAPYITDLEQAPRAGITRCPRLDYH KDDHLYAVTWFAAGAVLLAEVGVWGRSS" </pre>
gene	CDS	1282 a 2639 c 2619 g 1581 t	<pre> /transl_table=11 /evidence=experimental /product="putative heme A synthase or assembly protein" /protein_id="CAB56822.1" /db_xref="GI:6006416" /translat="MDHEPTIARDSDRTARKGVKGLGRDVLVASICGVALLMVGAS YAAPFVMEKCRATGENTGTQVATAAPGTGLAKIVREDSNADPLPMKFEPESE LEVINIGQVTFVTVINOAAPRTAGCAAVNAPLTVGSYFQKINCEFTFOTAPAEK RMDVAVYNDPSIADHDENDGLSTITLSVTFYPAKDPVYKPLASGEGRKKNL" 5981..6877 /EC_number="1.9.3.1" /function="cytochrome c oxidation (respiration)" /codon_start=1 /transl_table=11 /product="cytochrome c oxidase subunit III" /protein_id="CAB56823.1" /db_xref="GI:6006417" /translat="MTATAGCKNHIDYILVDPSPWPAVGSLSAFINAVGALMMHMEFA APVAYGATFGVLTVMASWMDGVYIEQAKYGDHTFVVDLHNRYSMLITLSEVVFVA WPAVFNAAALFPRADAVNATPRADAVNATPRADAVNATPRADAVNATPRADAVN TLLTLLTGTTVVAHHALLLENDROGKRYKLIITVVLGALFTVCQAVAEVSHAASFEAGN LYGATFEPMATGFGFHVLTGTTFLVLCFLRAVAGHFTKQHLGFEFAWYIYHFPDYW LFLFLCYVWGHGCAETMAHGAH" 6975..7376 /EC_number="1.9.3.1" /function="cytochrome c oxidase subunit III" /protein_id="CAB56824.1" /db_xref="GI:6006418" /translat="MNDTGAPEPKSETPTVLQSMARGLACPCRGQGLYAGFLTLAASS CDRGDVAFLDIDGDPALFIMLAGIIVACALIVYKQPPYLLAHLVWLPLIPLAT TLPLPRAMKSLILALQPHKAAAGRLVDRAK" 7199..8121 /EC_number="1.9.3.1" /function="cytochrome a3 biogenesis" /codon_start=1 /transl_table=11 /product="SUR1 homolog" /protein_id="CAB56825.1" /db_xref="GI:6006419" /translat="MRADRGQPPAAVLAACRAVVAADBRHNAALATARDEVAADRAAI PROGAGPAGRPRENNETARKPRVAGFALTLLTAFAALGWQJORTAKHEILIAE LTERIAAPIALPPAPMAALNPRADEFREVSFTAPASPDMAVYSSGSAVRKAS POTNAFLPARLPSEGVYIDAGFVEMTMDORSDEDAVKKIYGVQVALTGILRFPEE PQMLTAPASRDKRLMFVFRVDYALIASLIGKGTVAAPYITDLEQAPRAGITRCPRLDYH KDDHLYAVTWFAAGAVLLAEVGVWGRSS" </pre>
gene	CDS	1282 a 2639 c 2619 g 1581 t	<pre> /transl_table=11 /evidence=experimental /product="putative heme A synthase or assembly protein" /protein_id="CAB56822.1" /db_xref="GI:6006416" /translat="MDHEPTIARDSDRTARKGVKGLGRDVLVASICGVALLMVGAS YAAPFVMEKCRATGENTGTQVATAAPGTGLAKIVREDSNADPLPMKFEPESE LEVINIGQVTFVTVINOAAPRTAGCAAVNAPLTVGSYFQKINCEFTFOTAPAEK RMDVAVYNDPSIADHDENDGLSTITLSVTFYPAKDPVYKPLASGEGRKKNL" 5981..6877 /EC_number="1.9.3.1" /function="cytochrome c oxidation (respiration)" /codon_start=1 /transl_table=11 /product="cytochrome c oxidase subunit III" /protein_id="CAB56823.1" /db_xref="GI:6006417" /translat="MTATAGCKNHIDYILVDPSPWPAVGSLSAFINAVGALMMHMEFA APVAYGATFGVLTVMASWMDGVYIEQAKYGDHTFVVDLHNRYSMLITLSEVVFVA WPAV</pre>

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
U33883	2	GI:8708897		
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PubMed				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				


```

/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAG03497.1"
/db_xref="GI:9945934"
/translation="MSDAKVTIRLVGRLLVTVLMFAFGALVLYVMCRALGJNG
KTSASVSGDOVDVGRVGVQFMTSNINIDMWERFSGDQLVHPGAVNQMFYARN
RHTVLATLFDIDARKRPVAVGR"
7028..7915
/gene="coi1"
/feature="coi1"
/feature="PA0108"
7028..7915
/gene="coi1"
/codon_start=1
/product="cytochrome c oxidase, subunit III"
/protein_id="AAG03498.1"
/db_xref="GI:9945935"
/translation="MASHENYVPAQSKWPIIASIGLVTVFGLTGMENDLTAHKE
HGMVFEVGLITIAMLFQFQNVIRKSNGLVSAQMDRSFQMGMSWPLFSEVMEFA
FPGALFVYRHRAGMVGEGAKGVANMLNMFQYSPMLTQTPDKLPPPSAIEFWK
ELGTLGAGITGTFEMLTGFGAHVTLGALLIGIMLRIRLGRHDFAEHFGFEASW
YMHVDVWICLFTFYVI"
complement(7931..8140)
/gene="PA0109"
complement(7931..8140)
/gene="PA0109"
/codon_start=1
/translation="hypothetical protein"
/protein_id="AAG03499.1"
/db_xref="GI:9945936"
/translation="MLKVALVLLATLVLSFSGFLVKQDGHGSRVNSLVRYVL
AAATLVVAMGFESELSHAPWH"
8156..8950
/gene="PA0110"
8156..8950
/gene="PA0110"
/codon_start=1
/translation="hypothetical protein"
/protein_id="AAG03500.1"
/db_xref="GI:9945937"
/translation="MGFMQSORLSEAGNDRSRKGAFRGLPTLVVGLLPLVL
WLGWQIARADEKRALLSYFARGAPVSPGQELRDPAYVRLHGRFPERHTLL
YLPPGGLHLDADAGMPRLVTDLPALMQAFGRAGLPMELRLEPGDASFDTWPL
VSMPPERTGYAVQWFLATLALTYLVIGVRARKNHESRDSA"
8925..9503
/gene="PA0111"
8925..9503
/gene="PA0111"
/codon_start=1
/translation="hypothetical protein"
/protein_id="AAG03501.1"
/db_xref="GI:9945938"
/translation="MSLAIPMPERRRARGRLIIGLVGPMLLASAMYWNFW
LGRHSRAGALMGQTPADLVGQSGQEQMOLLVTAPGCGECCOOLVLAQINIA
PHENLVLRDADANAGKLLKDVQLLKSHIG"
9568..10641
/gene="PA0112"
9568..10641
/gene="PA0112"
/codon_start=1
/translation="hypothetical protein"
/protein_id="AAG03502.1"
/db_xref="GI:9945939"
/translation="MSGATRKPGFTVAVALTLAAVYVLLGAVTRLTAGLCPDWP

```

```

Query Match
Best Local Similarity 28.5%; Score 37.6; DB 1; Length 12829;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

CYGFVHVPLSEQLAHAEHPDPAVEAKQGMNEMIRYFAGALLIGLALHALVR
RGSDQPLKLLAVIAOAFGMVTVTKIMPOVYTAHLIGGFTTALLFLALR
LGRPAARRYPATRGLAGLLALLVIOGIALGQWSSNAVAACIDLPCHGEWPMRM

QY 2 cgcctgaacatcgttcgaagatgttcgcgtgtgcatcgttcacatgcgttcctc 61
Db 3976 CGGTGGAATTCCTCTGGAGCGGTGCGCTTCGTGATCTGCTGATGCGTCCCG 4035
QY 62 cgcctgcggtgctgttcacacagacagatcccccagaggtgcacatgcgttcac 121
Db 4036 CCACCCGACACCTGATCCACATCTACGACACTTCGACCGGAGCTGACGTCAC 4095
QY 122 aggg 125
Db 4096 CCGG 4099

RESULT 10
DMU36762/c
LOCUS
DEFINITION
DMU36762
ACCESSION
DHR38 (DHR38) mRNA, partial cds.
VERSION
U36762.1
KEYWORDS
GI:1036836
SOURCE
fruit fly strain-Canton S.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
Fisk, G.J. and Thummel, C.S.
Isolation, regulation, and DNA-binding properties of three
Drosophila nuclear hormone receptor superfamily members
Proc. Natl. Acad. Sci. U.S.A. 92 (23), 10604-10608 (1995)
2 (bases 1 to 1809)
Fisk, G.J. and Thummel, C.S.
Direct Submission
Submitted (20-SEP-1995) Gregory J. Fisk, Dept. of Human Genetics,
University of Utah, 5200 Eccles Institute of Human Genetics, Salt
Lake City, UT 84112, USA
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..1809
location/Qualifiers
/organism="Drosophila melanogaster"
/strain="Canton S"
/db_xref="taxon:7227"
/map="cytological map position 38L1.2"
1..1643
/gene="DHR38"
<1..1643
/gene="DHR38"
/note="nuclear hormone receptor superfamily member"
/codon_start=3
/product="DHR38"
/protein_id="AAC46926.1"
/db_xref="GI:1036837"
/translation="PPPLSGGWSASPPAPSOLOLHTLOSQAOQSHPNSSNNSSNNG
NSHNNGGYNHGHFNAINASNLSPSSASLSEYNGSADNFYCGOQOQOQOQOQOQOQO
GSTRASPKLAKITLNRHSHAHALJOLNSAPNSAASPSADLOAGRLQAFQSOICA
CLVGMWKEVVRDLSKGRGLPSPKSPESPSPISLTALVRSVHTDTPDSC
LDYSHYEDSMEADKVOQFOYLLTSSVDYIKQAEIIPYFDLLEDEDELFOASPL
FACIALTLFIRHGRIRPKVYEDLOKMIIGSLRDHYTNAEAKQKHQHFNSLGRKP
ELRSLVQGLQRIPLKLEDLVPAVALIEMVFTTLP"
BASE COUNT
422 a
587 c
479 g
321 t
ORIGIN

```



```

CDS
/gene="DHR38"
/locus="163..1818"
/codon_start=1
/protein_id="CAA61534.1"
/db_xref="GI:929562"
/db_xref="FLYBASE:FBgn0014859"
/db_xref="SWISS-PROT:P49869"
/translation="MDEDCCFPLSGGMSASPPAPSOLOQLHTLQSOAQMSPNNSNS
OOSVOOHYNHSHNERSLPTPTTSLAALTAVEAAATVGGPPVRAALPYOR
TVLRAGSTQSPKLAITLNRSHAHALJOLNSAPNSAASPASADLOAGRLQAP
SOLCAVCGDIACQHYGVTRCEGCGFFKRTVOKSGKYVLADKNCVDKRRNRGCF
CRFQCLVAVGWKEVVRVTRDILKGRGRRLPSKPSPOSPSPISLTALVSRVIT
PDSCLDYSHYEQSMSEADKVOEYOLTSVDYIKOPAEKIPGIFDLPEDELLF
QSASLEFLVRLAYRARDIPKILFCNGTATVLRHTQCRSGEMLNDIMEFSLNHLF
IDISAPACIALTLITERHGLRPKRYEOLMITSLRHRYNAEAKOKHYESRL
LGRPLRSLSVQGLQRIFYKLLEDLVPAFLIENMFVTLFP"
/misc_feature
/gene="DHR38"
/note="DNA-binding domain"
BASE COUNT      1594 a      1266 c      1061 g      1285 t
ORIGIN

Query Match      27.4%   Score 36.2: DB 3: Length 5206;
Best Local Similarity 57.5%: Pred. No. 6.6;
Matches 65: Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 20 cgattgtccggttgatgtcctcattcgtgctcgtcgcgcgggtgcttca 79
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 509 CCATTGTCGCGAGTGTGATGTTGATGATGATGATGATGATGATGATGATGAT 450
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 80 aacagcaagagtcgccgagggagacatcgatcaacgctgagaggtctagc 132
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 449 CCGTAGAAGTTCGCTGCGCAACACCATTTATTCGTAGAGGAACTGCG 397
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 13
LOCUS      SYNCTCONS      1260 bp      DNA      SYN      12-JUN-1993
DEFINITION Bacterial cytochrome oxidase subunit II consensus sequence, 5' end.
ACCESSION  L07491
VERSION     L07491.1 GI:295264
KEYWORDS   cytochrome oxidase II.
SOURCE      Artificial gene DNA.
ORGANISM   synthetic construct
REFERENCE   1 (bases 1 to 1260)
AUTHORS    Arraondo-Peter, R., Escamilla, E., Helman-Meneguzzi, F. and
            Hernandez, F.
TITLE       Computer analysis and modelling of a consensus sequence of
            bacterial cytochrome oxidase subunit II
JOURNAL     Unpublished (1992)
FEATURES
    source      1..1260
                Location/Qualifiers
                /organism="synthetic construct"
                /db_xref="taxon:32630"
                1..1260
                /gene="coxIcs"
                1..>1260
                /gene="coxIcs"
                /EC_number="1.9.3.1"
                /note="This sequence describes the consensus features of
                the previously published bacterial cytochrome oxidase
                subunit II. Y residues denotes hypervariable nucleotides.;
                bacterial consensus sequence; putative"
                /codon_start=1
                /transl_table=11
                /function="terminal oxidase"
                /label="COXIIS"
                /product="cytochrome oxidase II"
                /protein_id="AAA72326.1"

CDS
/gene="DHR38"
/locus="163..1818"
/codon_start=1
/protein_id="CAA61534.1"
/db_xref="GI:929562"
/db_xref="FLYBASE:FBgn0014859"
/db_xref="SWISS-PROT:P49869"
/translation="MDEDCCFPLSGGMSASPPAPSOLOQLHTLQSOAQMSPNNSNS
OOSVOOHYNHSHNERSLPTPTTSLAALTAVEAAATVGGPPVRAALPYOR
TVLRAGSTQSPKLAITLNRSHAHALJOLNSAPNSAASPASADLOAGRLQAP
SOLCAVCGDIACQHYGVTRCEGCGFFKRTVOKSGKYVLADKNCVDKRRNRGCF
CRFQCLVAVGWKEVVRVTRDILKGRGRRLPSKPSPOSPSPISLTALVSRVIT
PDSCLDYSHYEQSMSEADKVOEYOLTSVDYIKOPAEKIPGIFDLPEDELLF
QSASLEFLVRLAYRARDIPKILFCNGTATVLRHTQCRSGEMLNDIMEFSLNHLF
IDISAPACIALTLITERHGLRPKRYEOLMITSLRHRYNAEAKOKHYESRL
LGRPLRSLSVQGLQRIFYKLLEDLVPAFLIENMFVTLFP"
/misc_feature
/gene="DHR38"
/note="DNA-binding domain"
BASE COUNT      1594 a      1266 c      1061 g      1285 t
ORIGIN

Query Match      27.4%   Score 36.2: DB 3: Length 5206;
Best Local Similarity 57.5%: Pred. No. 6.6;
Matches 65: Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 20 cgattgtccggttgatgtcctcattcgtgctcgtcgcgcgggtgcttca 79
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 509 CCATTGTCGCGAGTGTGATGTTGATGATGATGATGATGATGATGATGATGAT 450
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 80 aacagcaagagtcgccgagggagacatcgatcaacgctgagaggtctagc 132
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 449 CCGTAGAAGTTCGCTGCGCAACACCATTTATTCGTAGAGGAACTGCG 397
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 14
LOCUS      AB047553      16537 bp      DNA      circular      VRT      14-AUG-2001
DEFINITION Pleocoglossus altivelis mitochondrial DNA, complete genome.
ACCESSION  AB047553
VERSION     AB047553.1 GI:13537245
KEYWORDS   Pleocoglossus altivelis mitochondrial DNA.
SOURCE      Mitochondrion Pleocoglossus altivelis
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Osmeriformes; Pleocoglossidae; Pleocoglossus.
REFERENCE   1 (bases 1 to 16537)
AUTHORS    Ishiguro, N., Miya, M. and Nishida, M.
TITLE       Complete Mitochondrial DNA Sequence of ayu Pleocoglossus altivelis
JOURNAL     Fish. Sci. 67, 474-481 (2001)
REFERENCE   2 (bases 1 to 16537)
AUTHORS    Ishiguro, N., Miya, M. and Nishida, M.
TITLE       Direct Submission
JOURNAL     Submitted (22-AUG-2000) Naoya Ishiguro, Ocean Research Institute,
            University of Tokyo, Marine Life Science; Minamidai 1-15-1, Nakano,
            Tokyo 164-8639, Japan (E-mail:ishiguro@ori.u-tokyo.ac.jp,
            Tel:81-3-5351-6487, Fax:81-3-5351-6488)
FEATURES
    source      1..16537
                Location/Qualifiers
                /organism="Pleocoglossus altivelis"
                /organella="mitochondrion"
                /db_xref="taxon:61084"
                1..68
                /product="tRNA-Phe"
                69..1013
                /gene="12S rRNA"
                /note="putative"
                /product="12S ribosomal RNA"
                69..1013
                /gene="12S rRNA"
                1014..1084
                /product="tRNA-Val"
                1085..2797
                /gene="16S rRNA"
                /note="putative"
                /product="16S ribosomal RNA"
                1085..2797
                /gene="16S rRNA"
                2798..2871
                /note="codon: UUR"
                /product="tRNA-Leu"

```

```

gene      2872..3846
/ gene="ND1"
CDS       2872..3846
/ gene="ND1"
/ codon_start=1
/ transl_table=2
/ product="NADH dehydrogenase subunit 1"
/ protein_id="BAB40601.1"
/ db_xref="GI:13537246"
/ translation="MITTILTHVINPLATVPLVAVLFLTLERVLGYMOLKCPN
ITGVYGLIPLADGKLFKEIPRSPSPFLPLPMLTLTALTMAMPPIYPA
DLNIGLIFLAPLSSLAAYSLGSGMANSKALIGALRAVQTIYSVSLGILISII
IFSGFTLOTENTQESIWLLPAMPPLAAMWYSLIAETRAPEDLTEGSESLVSGFN
VEYAGGPALFEFLAEVANILMTLSLVLELGSASHIPSLPCLTACNLMTEKAAALSVF
LWVRASYPFRFYDQIMHLWKNFLPLTALVLMHLALPLAFGLPQL"
tRNA      3847..3918
/ product="tRNA-Ile"
/ complement(3918..3988)
tRNA      3988..4056
/ product="tRNA-Gln"
/ protein_id="BAB40602.1"
/ db_xref="GI:13537247"
/ translation="MNPYVTSILISALIGTALTFSSSHMLAMGLEINTLAIPLM
TQHSRPSVEAAKRYITQAAATIMFASTNAMLTGSDIOLSHPAATIAFMAL
ALVYGLAPVHFWLPEVIGLITGILSTNOKIAPALIOITPALNSVLYCGLA
STLIGSGMNNOTQIRKILAYSSIAHGMWITISVSOYSPSLAICITLYIVMTSAFL
MKATAASLNTLTATMAKTPALATLALSLSLGGLPUSGTPMKHLLIOEMTKGCLP
LANTLAITASLFFYLRYCYAMTLTISPMLGASPSWRFSSNOSLLALSTVGT
GLIPTEFLASFSMA"
tRNA      5112..5183
/ product="tRNA-Trp"
/ complement(5185..5253)
tRNA      /product="tRNA-Ala"
tRNA      /complement(5255..5327)
tRNA      /product="tRNA-Asn"
/ complement(5356..5421)
tRNA      /product="tRNA-Cys"
/ complement(5421..5489)
tRNA      /product="tRNA-Tyr"
gene      5491..7041
/ gene="CO1"
/ note="CO1"
CDS       5491..7041
/ gene="CO1"
/ codon_start=1
/ transl_table=2
/ product="cytochrome oxidase subunit 1"
/ protein_id="BAB40603.1"
/ db_xref="GI:13537248"
/ translation="MAITRWEFSTNHKDIGTLYLIFEGMAGNVTALSLITRAFLSOP
GALLDDOIYVIVTAHAFVMEFVWMPIMIGFGNWLIPMIGAPMAPPFRANMSE
WLLPSPILLASSGVEAGAGTGVVYPLAGNLAHAGASVDLTFSLHLGSSILG
AINEFTTITNMKPAISIOYQPLFEWAVLITAVALLSLPVLAAIGITMLIDRNLTNT
FTDPAGGDPITIXOHLFWPFGCHPEVYITILPGFGMISHIVAYSGKKEPFGMOMVA
MMAIGLIGFIVAHMHFTYGMNDVPRAYETSATMTIAIPGVKFSMLATLHGSIIAM
ETPLMALGFIPLFTVGGILGIVLANSSLDLYHDTTYVVAHPIYVLSMGVFAIILA
FVHMPLEFSGYTLHSTWPKIHFGIMEVGNVLTFFQHLGLAGMPRRYSDDPYATTM
NTVSSIGSLISLVAVIMPLFLLMEAFAAKRENVSEVLSTVENVLHGCPPIYHFEER
AFVOYVAN"
/ complement(7042..7112)
tRNA      /note="codon: UCN"
tRNA      /product="tRNA-Ser"
tRNA      7117..7188
/ product="tRNA-Asp"
/ gene="CO2"
CDS       7203..7893
/ gene="CO2"
/ note="CO2"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/ codon_start=1
/ transl_except="(pos:7893,aa:TERM)"
/ transl_table=2
/ product="cytochrome oxidase subunit 2"
/ protein_id="BAB40604.1"
/ db_xref="GI:13537249"
/ translation="MAHPSQLGPDAAAPMEELHFDHAIWVLLJSTLVLYITVA
WVSTRKTKYLLDSQDEIEITVYPAVILILILPLSLRLYLMDEINDPILIKAMGH
QWYSYETDYDEIGFDSYKIPYDOLIPGOFRLLEDHNRVVESEPIRVLSAEDVL
HSMVAPALGVKMDVAPGLNOTAFIASRPGVIGGCSEIGCANHSMPIVEAVPLKH
FENMSMMLEDA"
tRNA      7894..7968
/ product="tRNA-Lys"
tRNA      7970..8137
/ gene="Atpase8"
CDS       7970..8137
/ gene="Atpase8"
/ codon_start=1
/ transl_table=2
/ product="adenosine triphosphatase subunit 8"
/ protein_id="BAB40605.1"
/ db_xref="GI:13537250"
/ translation="MPQLNPAFWALIVFSLWFLVTVIPKVLGVFSENEPTVQSAEK
TNPESNMPWH"
tRNA      8128..8810
/ gene="Atpase6"
CDS       8128..8810
/ gene="Atpase6"
/ note="Atpase6"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/ codon_start=1
/ transl_except="(pos:8809..8810,aa:TERM)"
/ transl_table=2
/ product="adenosine triphosphatase subunit 6"
/ protein_id="BAB40606.1"
/ db_xref="GI:13537251"
/ translation="MPLSPDOPMSVPPLGVLALALSLPILPPTFARMNRL
TLQGFVTRFTQULPLPVGKHKATITLSMLFLITVMGLPLPYFTPTQLSLN
MGLAFLPMLATVYIIGMRNOPALGHLBPBGTPVPIPVLIITISLIRPLALGVR
LTANLTAGHLLIOLIAATPAFVLLPLMPTVALITSAYFLTLLEAVAMIGAVPVLL
MSLYLOENV"
tRNA      8811..9595
/ gene="CO3"
CDS       8811..9595
/ gene="CO3"
/ note="CO3"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/ codon_start=1
/ transl_except="(pos:9594..9595,aa:TERM)"
/ transl_table=2
/ product="cytochrome oxidase subunit 3"
/ protein_id="BAB40607.1"
/ db_xref="GI:13537252"
/ translation="MAHQAHAFFHWDSPMPPLTGAVALLMTSGLAIMEFHSVILAV
AGTATLLIMYOMWRDIVERSTFGCHHTPVOKGLRYGMILFTSEVFFLGFMAY
HSLAPTEIGECWMPPTGITLUDPEVPLMTVAVLASGYVTVAHNSIMEGRKOTI
HSLTLLILGFEFTPLGLEVEYEAFTIADGVSTFVATGFRGLHVIIGSTPLALC
ILROIQYHTSSEHHGFEFAAMVYHFDVVMFLFYVSLYMWGS"
tRNA      9596..9666
/ product="tRNA-Gly"
tRNA      9667..10015
/ gene="ND3"
CDS       9667..10015
/ gene="ND3"
/ note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"

```


Thu Dec 27 08:21:40 2001

us-09-712-768-3.rge

Title: US-09-712-768-3
 Perfect score: 132
 ...tactctttagc ...tcaacgtctcagggctcgttagc 132

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

```

```

post-processing: Minimum Match 0%
                  Maximum Match 100%

```

Listing first 45 summaries

```
Database : N_Geneseq_1101:*
1: /SIP8/qcqd/data/geneseq/geneseqn/NA1980.DAT:*
1: /SIP8/qcqd/data/geneseq/geneseqn/NA1981.DAT:*
```

- 1: /SID58/gcgdata/geneseq/geneseqn/NA1980.DAT.*
- 2: /SID58/gcgdata/geneseq/geneseqn/NA1981.DAT.*
- 3: /SID58/gcgdata/geneseq/geneseqn/NA1982.DAT.*
- 4: /SID58/gcgdata/geneseq/geneseqn/NA1983.DAT.*
- 5: /SID58/gcgdata/geneseq/geneseqn/NA1984.DAT.*
- 6: /SID58/gcgdata/geneseq/geneseqn/NA1985.DAT.*
- 7: /SID58/gcgdata/geneseq/geneseqn/NA1986.DAT.*
- 8: /SID58/gcgdata/geneseq/geneseqn/NA1987.DAT.*
- 9: /SID58/gcgdata/geneseq/geneseqn/NA1988.DAT.*
- 10: /SID58/gcgdata/geneseq/geneseqn/NA1990.DAT.*
- 11: /SID58/gcgdata/geneseq/geneseqn/NA1991.DAT.*
- 12: /SID58/gcgdata/geneseq/geneseqn/NA1992.DAT.*
- 13: /SID58/gcgdata/geneseq/geneseqn/NA1993.DAT.*
- 14: /SID58/gcgdata/geneseq/geneseqn/NA1994.DAT.*
- 15: /SID58/gcgdata/geneseq/geneseqn/NA1995.DAT.*
- 16: /SID58/gcgdata/geneseq/geneseqn/NA1996.DAT.*
- 17: /SID58/gcgdata/geneseq/geneseqn/NA1997.DAT.*
- 18: /SID58/gcgdata/geneseq/geneseqn/NA1998.DAT.*
- 19: /SID58/gcgdata/geneseq/geneseqn/NA1999.DAT.*
- 20: /SID58/gcgdata/geneseq/geneseqn/NA2000.DAT.*
- 21: /SID58/gcgdata/geneseq/geneseqn/NA2001.DAT.*
- 22: /SID58/gcgdata/geneseq/geneseqn/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	132	100.0	132	22	AAA91490	G. oxydans cytochr
2	32.2	24.4	9810	20	AAZ32025	Human MERT1 relate
3	32.2	24.4	9810	22	AF018073	CDNA clone
4	32	24.2	182	22	AAAG90082	Human lung tumour
5	32	24.2	195	21	AAAF68838	CDNA encoding huma
6	32	24.2	195	22	AAI28839	Colon tumour relat
7	32	24.2	230	22	AAF68647	Human lung tumour
8	32	24.2	236	21	AAA78100	CDNA encoding huma
9	32	24.2	236	22	AAI28838	Colon tumour relat
10	32	24.2	237	22	AAF68634	Human lung tumour
11	32	24.2	251	21	AAZ80445	Human colon cancer

12	32	24.2	270	22	AAF68700	human
13	32	24.2	272	22	AAI29219	Colon tumour relat
14	32	24.2	275	22	AAF66858	Human lung tumour
15	32	24.2	275	22	AAF66808	Human lung tumour
16	32	24.2	277	22	AAF66518	Human lung tumour
17	32	24.2	281	22	AAF68779	Human lung tumour
18	32	24.2	288	22	AAF66670	Human lung tumour
19	32	24.2	289	22	AAF66656	Human lung tumour
20	32	24.2	294	22	AAF66646	Human lung tumour
21	32	24.2	306	22	AAF68546	Human lung tumour
22	32	24.2	313	21	AAF68099	CDNA encoding huma
23	32	24.2	313	22	AAI28837	Colon tumour relat
24	32	24.2	353	21	AAA16405	Human colon cancer
25	32	24.2	513	20	AAV87034	EST clone BG95. H
26	32	24.2	520	21	AAV78262	Human cancer assoc
27	32	24.2	535	22	AAH35343	Human colon cancer
28	32	24.2	563	22	AAH35344	Human colon cancer
29	32	24.2	595	21	AAH38353	Human cancer assoc
30	32	24.2	682	21	AAV87821	Human cancer assoc
31	32	24.2	708	19	AAV43711	Cancer associated
32	32	24.2	747	22	AAH35342	Human colon cancer
33	32	24.2	789	21	AAA16534	Human colon cancer
34	32	24.2	854	16	AAV05906	Mitochondrial cytol
35	32	24.2	854	22	AAF66880	Mitochondrial cytol
36	32	24.2	2667	22	AAH14061	Human CDNA sequenc
37	32	24.2	16569	20	AAZ11492	Mutated human mlti
38	32	24.2	16569	20	AAZ11493	Mutated human mlti
39	32	24.2	16569	21	AAZ57203	Human mitochondria
40	32	24.2	16569	21	AAZ57203	Human mitochondria
41	31.6	23.9	253	22	AAF66697	Human lung tumour
42	31.6	23.9	559	21	AAV73873	Human secreted pr
43	31	23.5	173	22	AAF68596	Human lung tumour
44	31	23.5	283	22	AAF68559	Human lung tumour
45	31	23.5	366	22	AAF68607	Human lung tumour

ALIGNMENTS

RESULT	1
XX	AAA91490
XX	ID AAA91490 standard; DNA; 132 BP.
XX	AC
XX	AAA91490;
XX	06-AUG-2001 (first entry)
XX	G. oxydans cytochrome C oxidase (COII) coding sequence fragment.
XX	DE
XX	Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;
XX	KW oxidative fermentation; electron transfer; respiratory chain; L-sorbose
XX	KW 2-Keto-L-gluconic acid; 2KGa production; aldehyde production;
KW	carboxylic acid production; ketone production; ds.
XX	OS Gluconobacter oxydans.
XX	XX
FH	Key Location/Qualifiers
FT	1..132
FT	/*tag= a
FT	/product= "COII"
XX	EP103603-A2.
XX	PH
XX	30-MAY-2001.
XX	PD
XX	14-NOV-2000; 2000EP-0124785.
XX	PF
XX	17-NOV-1999; 99EP-0122842.
XX	PR
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.
PA	
XX	
XX	Asakura A, Hoshino T, Shinjoh M;
PI	
XX	

DR WPI: 2001-357953/38.
P-PSDB: AAY97751.

XX New cytochrome c oxidase complex having cytochrome c oxidase activity
PT from Gluconobacter oxydans DSM 4025, useful in mediating electron
PT transfer in respiratory chain or producing 2-keto-L-gulonate acid from
XX L-sorbose or D-sorbitol

PS Claim 14: Page 24: 42pp: English.

XX This sequence encodes a fragment of the Gluconobacter oxydans
CC cytochrome c oxidase, COI. The invention relates to a cytochrome c
CC oxidase COI complex. The COI complex is useful in improving oxidative
CC fermentation and is an essential component mediating electron transfer
CC in the respiratory chain. The recombinant microorganism and the
CC cytochrome c oxidase may be used in the genetic preparation of a
CC recombinant COI complex and in the production of 2-keto-L-gulonate acid
CC (2KG) from L-sorbose or D-sorbitol in a culture medium. The COI is also
CC useful as a terminal oxidase, oxidizing cytochrome c, an electron
CC acceptor from an enzyme belonging to dehydrogenase for the production of
CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,
CC especially the production of 2KG from L-sorbose or D-sorbitol.

XX Sequence 132 BP: 22 A; 34 C; 40 G; 36 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 132; DB 22; Length 132;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccgcgtgaatcgtcgtggaagatgttcgcgtgtgattcgtgttcattcgtgcgttc 60
Db 1 ccgcgtgaatcgtcgtggaagatgttcgcgtgtgattcgtgttcattcgtgcgttc 60
QY 61 tcctgcgcgtgtcgttcacacagaaggtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 60
Db 61 tcctgcgcgtgtcgttcacacagaaggtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
QY 121 gaggtgcgtgac 132
Db 121 gaggtgcgtgac 132

RESULT 2
AA232025
ID AA232025 standard; DNA; 9810 BP.
XX
AC AA232025;

DT 10-JAN-2000 (first entry)

DE Human METH1 related EST AF018073.

XX Human: METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
KW angiogenesis inhibitor; abnormal wound healing; inflammation;
KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KW diabetic retinopathy; macula degeneration; haemangioma; detection;
XX arterial-venous malformation; immune deficiency; ss.

OS Homo sapiens.

XX WO9337660-A1.

XX 29-JUL-1999.

XX 22-JAN-1999; 99MO-USO1313.

XX 23-JAN-1998; 98US-0072298.

XX 28-AUG-1998; 98US-0098539.

XX (IRUE/) IRUELA-ARISPE L.

XX (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

XX IrueLA-Arispe L, Hastings GA, Ruben SM;

XX WPI: 1999-590684/50.

XX New isolated metalloprotease thrombospondin polypeptides, useful for
PT treating hyperproliferative disorders, cancers or autoimmune disorders
PT

PS Disclosure: Page 353-359; 457pp: English.

XX AA232000 and AA232001 encode, and AAY9501 and AAY9502 represent, human
CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC macula degeneration, haemangiomas, and arterial-venous malformations.
CC They may be useful in treating deficiencies or disorders of the immune
CC system, by activating or inhibiting the proliferation, differentiation,
CC or mobilisation (chemotaxis) of immune cells. The etiology of these
CC cancer or some autoimmune disorders may be genetic, somatic, such as
CC toxins, or infectious. They can also be used to treat inflammatory or
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AA232002 to AA232080, and AAY9503 to
CC invention.

XX Sequence 9810 BP: 1583 A; 3401 C; 3201 G; 1625 T; 0 other;

Query Match
Best Local Similarity 24.4%; Score 32.2; DB 20; Length 9810;
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 3 gctggaaatcgtcgtggaagatgttcgcgtgtgattcgtgttcattcgtgcgttc 62
Db 4608 gccgggcatcgtcgtcgaagcgttcctgaacggtgacctcgcctcgaagcagcgccttcg 4667
QY 63 gctggcgcgtcgttcacacagaaggtcc 95
Db 4668 gacgtcgcagcgtcgaacacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4700

RESULT 3
AAC90082
ID AAC90082 standard; DNA; 9810 BP.
XX
AC AAC90082;

DT 19-MAR-2001 (first entry)

DE AF018073 cDNA clone.

XX METH: metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; nonunion fracture;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischemic limb angiogenesis; Osler-Weber syndrome; wound granulation;
KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;
KW angiobroma; fibromuscular dysplasia; expressed sequence tag;
XX Crohn's disease; atherosclerosis; birth control; ss.

OS Unidentified.

XX WO200071577-A1.

XX 30-NOV-2000.

XX 25-MAY-2000; 2000WO-US1446Z.
 PF
 XX 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK-) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBEN/) RUBEN S M.
 PA (JONAK/) JONAK Z L.
 PA (TRULLI/) TRULLI S H.
 PA (FORN/) FORNWARD J A.
 PA (TERRETT/) TERRETT J A.
 XX
 PI IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornward JA, Terrett JA;
 DR WPI: 2001-025136/03.
 XX
 PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 PS Claim 7: Pages 653-659; 768pp: English.
 XX
 CC The present invention relates to human METH1 and METH2, (ME for
 CC metalloproteinase and TN for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulomatous, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 CC
 XX
 SQ Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;
 Query Match 24.4%; Score 32.2; DB 22; Length 9810;
 Best Local Similarity 59.1%; Pred. No. 1.7;
 Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
 QY 3 gctggaatcgtcgtgacatgttcggtgtgattcgttcacatcgtgcttc 62
 Db 4608 ggcgggacgtcgtcgtcgtcgttcgttcgttcgttcgttcgttc 62
 QY 63 gctgctgctgtcgttcacacagcaagattcc 95
 Db 4668 gacgtcgtcgttcacacacgtcgttcgttcgttcgttcgttc 4700
 RESULT 4
 AAF68838
 ID AAF68838 standard; CDNA: 182 BP.
 AC AAF68838;
 XX 12-APR-2001 (first entry)
 DT Human lung tumour protein related nucleotide sequence SEQ ID NO:773.
 XX

KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 KW cytostatic; antisense inhibition; ss.
 XX
 OS Homo sapiens.
 PN WO200100828-A2.
 PD 04-JAN-2001.
 XX
 PF 30-JUN-2000; 2000WO-US18061.
 XX
 PR 30-JUN-1999; 99US-0346492.
 PR 15-OCT-1999; 99US-0419356.
 PR 17-DEC-1999; 99US-0466867.
 PR 30-DEC-1999; 99US-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-0533077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MW, Mannion J;
 XX
 DR WPI: 2001-071488/08.
 XX
 PT Lung tumor-associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -
 PS Claim 4: Page 402; 436pp: English.
 XX
 CC The present invention describes immunogenic portions of lung tumour-
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.
 CC (I) have cytostatic activity and can be used in gene therapy, antisense
 CC inhibition and in vaccines. The NAs and the lung tumour-associated
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with their inappropriate expression,
 CC especially lung cancers. For example, the NAs may be administered to
 CC treat diseases by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of the protein by expressing inactive proteins
 CC or to supplement the activity of the patients own production of (I). Additionally,
 CC the NAs may be used to produce the lung-tumour associated protein, according
 CC to standard recombinant DNA methodology. Conversely, antisense NA
 CC molecules may be administered to down regulate protein expression by
 CC binding with the cells own genes and preventing their expression. The NA
 CC and complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar NA sequences in
 CC samples, and hence which patients may be in need of treatment for lung
 CC cancer. The (I) may be used as antigens in the production of antibodies
 CC and in assays to identify modulators (agonists and antagonists) of the
 CC expression and activity of the protein. AAF68083 to AAF68878 and
 CC AAB76848 to AAB76878 represent human lung tumour protein related
 CC nucleotide and protein sequences which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 182 BP; 55 A; 65 C; 20 G; 42 T; 0 other;
 Query Match 24.2%; Score 32; DB 22; Length 182;
 Best Local Similarity 55.4%; Pred. No. 0.63;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 5 tgggaatcgtcgtgacatgttcggtgtgattcgttcacatcgtgcttcgc 64
 Db 62 tagaacgtcgtcgtcgtcgttcgttcgttcgttcgttcgttcgttc 121
 QY 65 tggcgtgctgttcacacagcaagattcccgaggaggtacatcgtcaca 116
 Db 122 tagcatccttaccataacagacaggtacacgttcctccctaccacaa 173

RESULT 5
 AAA78101
 ID AAA78101 standard; cDNA: 195 BP.
 AC
 AC AAA78101:
 DT 14-NOV-2000 (first entry)
 DE CDNA encoding human colon tumour polypeptide, SEQ ID NO:388.
 DE
 KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
 KW Immunotherapy; diagnosis; progression; ss.
 OS
 OS Homo sapiens.
 PN WO200037643-A2.
 PD 29-JUN-2000.
 PF 23-DEC-1999; 99WO-US30909.
 PR 23-DEC-1998; 98US-0221298.
 PR 02-JUL-1999; 99US-0347496.
 PR 22-SEP-1999; 99US-0401064.
 PR 19-NOV-1999; 99US-0444242.
 PR 02-DEC-1999; 99US-0454150.
 PA (CORI-) CORIXA CORP.
 PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk J;
 PI Wang T, Yugu J;
 DR WPI: 2000-442671/38.
 XX New colon tumor polypeptides used to inhibit the development of cancer,
 PT especially colon cancer, and for diagnosing and monitoring the
 PT progression of the cancer -
 PS
 PS Claim 1, Page 203; 229pp; English.
 CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
 CC portions of proteins which are associated with human colon tumours.
 CC The invention also specifically discloses 8 human colon tumour proteins
 CC (AA81897-B11904). The nucleic acids, the polypeptides they encode, and
 CC antigen presenting cells (APCs, preferably dendritic cells) expressing
 CC such polypeptides may be used in vaccines that target tumour cells,
 CC especially colon tumour cells, thereby inhibiting the development of
 CC cancer. T-cells specific for the polypeptide expressed by the APC are
 CC used to remove tumour cells from biological samples, especially blood or
 CC fractions thereof. The sample or the isolated T-cells specifically for the
 CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
 CC CD8+ T-cells from a patient may be incubated with a polypeptide or
 CC nucleic acid of the invention, or an APC expressing such a polypeptide,
 CC to cause the proliferation of specific T-cells. The T-cells can be
 CC cloned and then administered back to the patient to inhibit cancer
 CC development. Nucleic acids encoding the polypeptides and antibodies
 CC against the polypeptides may be used to determine the expression level
 CC of a tumour protein of the invention, and therefore to determine whether
 CC cancer cells are present. Such diagnostic methods may also be used to
 CC monitor the progression of a cancer by repeating the processes at time
 CC intervals, and comparing the current result to previous results. The
 CC present sequence represents a cDNA encoding a human colon tumour
 CC polypeptide.
 CC
 CC Sequence 195 BP: 59 A; 70 C; 22 G; 44 T; 0 other;

Query Match 24.2%; Score 32; DB 21; Length 195;
 Best Local Similarity 55.4%; Pred. No. 0.65;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtctgacgatgttcggtgtgattcgtgcttcacatcgatgcttcgcgc 64

Db 64 tagaaacgctcgaactacccctccgacatccatccatccatccatccatccatcc 123
 OY 65 tggcgtgtcgttcaaacagaagatcccccgggtgacatcgtatcaaa 116
 Db 124 taagcatccttacaatacagacgaggtcaacatccctccatccatcaaa 175
 RESULT 6
 AAI28839
 ID AAI28839 standard; cDNA: 195 BP.
 AC
 AC AAI28839;
 DT 12-OCT-2001 (first entry)
 DE Colon tumour related determined cDNA sequence for clone 31998.
 DE
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW Immunogenic; gene therapy; vaccine; colonic cancer; ss.
 OS
 OS Homo sapiens.
 PN WO200149716-A2.
 PD 12-JUL-2001.
 PF 29-DEC-2000; 2000WO-US3596.
 PR 30-DEC-1999; 99US-0476296.
 PR 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 PA (CORI-) CORIXA CORP.
 PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 DR WPI: 2001-441847/47.
 XX Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 PT
 PS Claim 2, Page 240; 472pp; English.
 CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
 CC and AAI24494 to AAI24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 195 BP; 59 A; 70 C; 22 G; 44 T; 0 other;
SQ

Query Match	24.28;	Score 32;	DB 22;	Length 195;
Best Local Similarity	55.4%;	Pred. No. 0.65;		
Matches 62;	Conservative	0;	Mismatches 50;	Indels 0;
			Gaps	0

QY 5 tggaaatcgtcggacgagttgttcggtttgatctcggctcctacggatgcgttcgc 64
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 64 tagaaacgcgtcgaactatcctgcgcgcacatccttgatctccatcgcgcctccatcc 123
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 65 tgcgcgtgctgttcaacacgaagaattccgcgaggtgacatcgtcatca 116
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 124 taacgacatccttacaataacagacgggttcaacgactccctcccttaccatcaa 175
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT	7
AAAF68647	
ID	AAAF68647 standard; cDNA; 230 BP.

AC	AAF68647;
XX	
DT	12-APR-2001 (first entry)

Human lung tumour related nucleotide sequence SEQ ID NO:582.

Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
cytostatic; antisense inhibition; ss.

OS	Homo sapiens.
XX	
PN	W0200100828-A2

PD 04-JAN-2001

AA
PF 30-JUN-2000; 2000WO-US18061

PR	30-JUN-1999:	99US-0346492
PR	15-OCT-1999:	99US-0419356
PR	17-DEC-1999:	99US-0468687
PR	30-DEC-1999:	99US-0476300
PR	06-MAR-2000:	2000US-0519642
PR	22-MAR-2000:	2000US-0533072
PR	10-APR-2000:	2000US-0546259
PR	27-APR-2000:	2000US-0560406
PR	05-JUN-2000:	2000US-0589184

PA (CORI-) CORIXA CORP

Wang T, Bangor CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D,
Retter MW, Mannion J;

DR WPI; 2001-071488/08.

AA Lung tumor-associated proteins and the nucleic acids that encode them
 PT useful for preventing, diagnosing and treating lung cancer -
 XX
 PS Claim 4; Page 344; 436pp; English.

AA The present invention describes immunogenic portions of lung tumour-
CC associated proteins (I) and the nucleic acids (NAs) that encode them.
CC (I) have cytostatic activity and can be used in gene therapy, antisense
CC inhibition and in vaccines. The NAs and the lung tumour-associated
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with their inappropriate expression,
CC especially lung cancers. For example, the NAs may be administered to
CC treat diseases by rectifying mutations or deletions in a patient's genome
CC that affect the activity of the protein by expressing inactive proteins
CC or to supplement the patients own production of (I). Additionally, the
CC NAs may be used to produce the lung-tumour associated protein, according
CC to standard recombinant DNA methodology. Conversely, antisense NA

molecules may be administered to down regulate protein expression by binding with the cells own genes and preventing their expression. The CC and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar NA sequences in CC samples, and hence which patients may be in need of treatment for lung cancer. The (I) may be used as antigens in the production of antipode cancer. and in assays to identify modulators (agonists and antagonists) of the expression and activity of the protein. AAF6803 to AAF6878 and AAF6648 to AAF6787 represent human Lung tumour protein related nucleotide and protein sequences which are used in the exemplification of the present invention.

Sequence 230 BP; 62 A; 79 C; 28 G; 61 T; 0 other;

Query Match	24.2%	Score 32;	DB 22;	length 230;
Best Local Similarity	55.4%;	Pred. No.	0.68;	
Matches	62;	Conservative	0;	Mismatches 50;
				Indels 0;
				Gaps 0.

[illegible]

RESULT	8
AAA78100	
ID	AAA78100 standard; cDNA; 236 BP

AC AAA78100;

DT 14-NOV-2000 (first entry)

XX
DE
cDNA encoding human colon tumour polypeptide, SEQ ID NO:387.

XX Human colon tumour polypeptide; tumour antigen; cancer; vaccine
KW immunotherapy; diagnosis; progression; ss.
KW

OS	Homo sapiens.
XX	
PN	W0200037643-A2

XX
PD 29-JUN-2000.

XX
XX
23-DEC-1999. 09W0-11S30909

PR	23-DEC-1998;	9805-0221298
PR	02-JUL-1999;	9905-0347496
PR	22-SEP-1999;	9905-0401064
PR	12-NOV-1999;	9905-0444242
PR	02-DEC-1999;	9905-0454150

PA (CORI-) CORIXA CORP

Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J, Wang T, Yudin J;

DR WPI; 2000-442671/38.

aa New colon tumor polypeptides used to inhibit the development of cancer
 pt especially colon cancer, and for diagnosing and monitoring the
 pt progression of the cancer -
 xx
 ps Claim 1; Page 203; 229pp; English.

AA Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or portions of proteins which are associated with human colon tumours. CC The invention also specifically discloses 8 human colon tumour proteins CC (The invention also specifically discloses 8 human colon tumour proteins CC (AAH1897-B11904). The nucleic acids, the polypeptides they encode, and CC antigen presenting cells (APCs, preferably dendritic cells) expressing

CC such polypeptides may be used in vaccines that target tumour cells,
 CC especially colon tumour cells, thereby inhibiting the development of
 CC cancer. T-cells specific for the polypeptide expressed by the APC are
 CC used to remove tumour cells from biological samples, especially blood or
 CC fractions thereof. The sample or the isolated T-cells specific for the
 CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
 CC CD8+ T-cells from a patient may be incubated with a polypeptide or
 CC nucleic acid of the invention, or an APC expressing such a polypeptide,
 CC to cause the proliferation of specific T-cells. The T-cells can be
 CC cloned and then administered back to the patient to inhibit cancer
 CC development. Nucleic acids encoding the polypeptides and antibodies
 CC against the polypeptides may be used to determine the expression level
 CC of a tumour protein of the invention, and therefore to determine whether
 CC cancer cells are present. Such diagnostic methods may also be used to
 CC monitor the progression of a cancer by repeating the processes at time
 CC intervals, and comparing the current result to previous results. The
 CC present sequence represents a cDNA encoding a human colon tumour
 CC polypeptide.

SQ Sequence 236 BP; 65 A; 83 C; 26 G; 62 T; 0 other;

Query Match 24.2%; Score 32; DB 21; Length 236;
 Best Local Similarity 55.4%; Pred. No. 0.68;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatgctcgagcagttcttcggtgtgattctgtcttcacgtcgtcttcgc 64
 DB 105 tagaaacgctcgtacatctcctgcgcacatcctagtcctccatccatccc 164
 OY 65 tgcgagtcgtctcaaacagcaagttcccccggaggtgacatcgtatcaaa 116
 DB 165 tagcatccttatacaataacagacgaggtcaacgacccctccatcaataa 216

RESULT 9

AAI28838
 ID AAI28838 standard; cDNA; 236 BP.

AAI28838;

12-OCT-2001 (first entry)

Colon tumour related determined cDNA sequence for clone 31934.

Human; Immunotherapy; diagnosis; colon cancer; colon tumour;
 Immunogenic; gene therapy; vaccine; colonic cancer; ss.

Homo sapiens.

WO200149716-A2.

12-JUL-2001.

29-DEC-2000; 2000WO-US35596.

30-DEC-1999; 99US-0476296.

10-JAN-2000; 2000US-0480321.

15-FEB-2000; 2000US-0504629.

06-MAR-2000; 2000US-0519444.

19-MAY-2000; 2000US-0575251.

29-JUN-2000; 2000US-0609448.

28-AUG-2000; 2000US-0649811.

(CORI-) CORIXA CORP.

Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y,

WPI; 2001-441847/47.

Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -

XX Claim 2; Page 240; 472pp; English.

PS The present invention describes colon tumour associated proteins (I) and
 XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
 XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
 XX (II) may be used in the prevention, diagnosis and treatment of diseases
 XX associated with inappropriate colon tumour associated protein (TCAP)
 XX expression, such as colonic cancer. For example, (I) and (II) may be
 XX used to treat disorders associated with decreased expression by
 XX rectifying mutations or deletions in a patient's genome that affect the
 XX activity of TCAPs by expressing inactive proteins or to supplement the
 XX patients own production of them. Additionally, (II) may be used to
 XX produce the TCAP proteins, by inserting the nucleic acids into a host
 XX cell culturing the cell to express the protein. (II) and its
 XX complementary sequences may also be used as DNA probes in diagnostic
 XX polymerase chain reaction (PCR) and hybridisation assays to detect and
 XX quantify the presence of similar nucleic acids in samples, and
 XX therefore which patients may be in need of restorative therapy. (I) may
 XX also be used as antigens in the production of antibodies against TCAPs
 XX and in assays to identify modulators of TCAP expression and activity.
 XX Anti-(I) antibodies and antagonists may also be used to down regulate
 XX TCAP expression and activity. The anti-(I) antibodies may also be used
 XX as diagnostic agents for detecting the presence of TCAPs in samples
 XX (e.g. by enzyme linked immunosorbent assay (ELISA)). AAI28460 to AAI29512
 XX and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
 XX given in the exemplification of the present invention.

SQ Sequence 236 BP; 65 A; 83 C; 26 G; 62 T; 0 other;

Query Match 24.2%; Score 32; DB 22; Length 236;
 Best Local Similarity 55.4%; Pred. No. 0.68;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatgctcgagcagttcttcggtgtgattctgtcttcacgtcgtcttcgc 64
 DB 105 tagaaacgctcgtacatctcctgcgcacatcctagtcctccatccatccc 164
 OY 65 tgcgagtcgtctcaaacagcaagttcccccggaggtgacatcgtatcaaa 116
 DB 165 tagcatccttatacaataacagacgaggtcaacgacccctccatcaataa 216

RESULT 10

AAF68634/C
 ID AAF68634 standard; cDNA; 237 BP.

AAF68634;

12-APR-2001 (first entry)

Human lung tumour protein related nucleotide sequence SEQ ID NO:569.

Human; Lung cancer; lung tumour; lung tumour protein; gene therapy;
 KM lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
 XX cytosolic; antisense inhibition; ss.

Homo sapiens.

WO200100828-A2.

04-JAN-2001.

30-JUN-2000; 2000WO-US18061.

30-JUN-1999; 99US-0346492.

15-OCT-1999; 99US-0419356.

17-DEC-1999; 99US-0468867.

30-DEC-1999; 99US-0476300.

06-MAR-2000; 2000US-0519642.

22-MAR-2000; 2000US-0533077.

10-APR-2000; 2000US-0546259.

PR 30-DEC-1999; 9905-0476300.
 PR 06-MAR-2000; 2000US-0519642.
 PR 22-MAR-2000; 2000US-053077.
 PR 10-APR-2000; 2000US-0546259.
 PR 27-APR-2000; 2000US-0560406.
 PR 05-JUN-2000; 2000US-0589184.
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Bangur GS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
 PI Retter MW, Mannion J;
 XX
 DR WPI; 2001-071488/08.
 XX
 PT Lung tumor-associated proteins and the nucleic acids that encode them,
 PT useful for preventing, diagnosing and treating lung cancer -
 XX
 PS Claim 4; Page 359; 436pp; English.
 CC The present invention describes immunogenic portions of lung tumour-
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.
 CC (I) have cytostatic activity and can be used in gene therapy, antisense
 CC inhibition and in vaccines. The NAs and the lung tumour-associated
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with their inappropriate expression,
 CC especially lung cancers. For example, the NAs may be administered to
 CC treat diseases by rectifying mutations or deletions in a patient's genome
 CC or to supplement the activity of the protein by expressing inactive proteins
 CC to standard recombinant DNA methodology. Conversely, antisense NA
 CC molecules may be administered to down regulate protein expression by
 CC binding with the cells own genes and preventing their expression. The NA
 CC and complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar NA sequences in
 CC samples, and hence which patients may be in need of treatment for lung
 CC cancer. The (I) may be used as antigens in the production of antibodies
 CC and in assays to identify modulators (agonists and antagonists) of the
 CC expression and activity of the protein. AAF68083 to AAF68878 and
 CC AAB76848 to AAB76878 represent human lung tumour protein related
 CC nucleotide and protein sequences which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 270 BP; 74 A; 30 C; 93 G; 73 T; 0 other;
 Query Match 24.2%; Score 32; DB 22; Length 270;
 Best Local Similarity 55.4%; Pred. No. 0.71;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 5 tggaaatcgtctggacgattgttcggtgtgattcgtcttcacgtgcttcgc 64
 DB 121 TAGAAGCCGCTGAACATCTGACCGCCGATCATCTGCTAGTCCCTCCATGCC 62
 QY 65 tggcgggtgctgttcaaacagcagaggttcccccggggtgacatcgtatcaaa 116
 DB 61 TACGCATCCTTTACATACAGACGAGGTCAACGATCCCTCCCTTACCATCAA 10
 RESULT 13
 AAI29219/c
 ID AAI29219 standard; cDNA; 272 BP.
 XX
 AC AAI29219;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Colon tumour related determined cDNA sequence for clone R0094:C02.
 XX
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer; ss.
 XX
 OS Homo sapiens.

XX
 PN W0200149716-A2.
 XX
 PD 12-JUL-2001.
 XX
 PF 29-DEC-2000; 2000WO-US355596.
 XX
 PR 30-DEC-1999; 9905-0476296.
 PR 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 XX
 DR WPI; 2001-441847/47.
 XX
 PT Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 XX
 PS Claim 2; Page 352; 472pp; English.
 CC The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AAI28460 to AAI29512
 CC given in the exemplification of the present invention.
 XX
 SQ Sequence 272 BP; 74 A; 30 C; 93 G; 75 T; 0 other;
 Query Match 24.2%; Score 32; DB 22; Length 272;
 Best Local Similarity 55.4%; Pred. No. 0.71;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 5 tggaaatcgtctggacgattgttcggtgtgattcgtcttcacgtgcttcgc 64
 DB 121 TAGAAGCCGCTGAACATCTGACCGCCGATCATCTGCTAGTCCCTCCATGCC 62
 QY 65 tggcgggtgctgttcaaacagcagaggttcccccggggtgacatcgtatcaaa 116
 DB 61 TACGCATCCTTTACATACAGACGAGGTCAACGATCCCTCCCTTACCATCAA 10
 RESULT 14
 AAF68758/c
 ID AAF68758 standard; cDNA; 275 BP.
 XX
 AC AAF68758;

XX Sequence 275 BP; 75 A; 95 C; 31 G; 73 T; 1 other;
SQ

Query Match 24.2%; Score 32; DB 22; Length 275;
Best Local Similarity 55.4%; Pred. No. 0.71;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tggaaatcgtctggacgattgttccggttgatctgttcatcgttggtctcgc 64
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 tagaaacggtctgaactatctctgcccgcacatcctagttcctcatcgccctccatccc 214
OY 65 tgcggtgtctgttcaacagcaagagttcccccagaggtgacatcgtcatcaa 116
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 tagcatccttatacataacagacgaggtcaacgaltccctcccttaccatcaa 266

Search completed: December 22, 2001, 10:44:26
Job time: 9212 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 09:36:14 ; Search time 113.93 Seconds
(Without alignments)
262.399 Million cell updates/sec

Title: US-09-712-768-3
Perfect score: 132
Sequence: 1 ccgcgtggaatcgtctgac.....tcaacgtcagaggtcgtacg 132

Scoring table: IDENTITY=NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCUTS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	24.2	251	4	US-09-328-111-529
2	32	24.2	353	4	US-09-385-982-410
3	32	24.2	754	1	US-08-219-842-2
4	32	24.2	754	1	US-08-451-096-2
5	32	24.2	789	4	US-09-385-982-539
6	32	24.2	854	4	US-08-413-740A-2
7	32	24.2	854	5	PCT-US95-04063-2
8	32	24.2	16569	4	US-09-087-889-1
9	31.2	23.6	6744	4	US-09-097-889-1
10	29.4	22.3	591	2	US-09-156-979-1
11	28	21.2	30001	1	US-08-125-468-1
12	28	21.2	30001	2	US-08-474-933-1
13	27.8	21.1	9551	1	US-08-056-200-93
14	27.8	21.1	9551	2	US-08-800-644-93
15	27.6	20.9	1600	2	US-08-776-210-3
16	27.6	20.8	571	2	US-08-803-899-1
17	27.2	20.6	783	2	US-08-505-218-9
18	27.2	20.6	1067	2	US-08-505-218-5
19	27.2	20.5	840	2	US-08-924-759-13
20	27	20.5	840	2	US-09-248-335-13
21	27	20.5	441529	4	US-09-103-840A-1
22	26.8	20.3	1557	5	PCT-US91-01327-11
23	26.8	20.3	1557	6	PCT-US91-01327-11
24	26.6	20.2	971	3	US-09-248-335-65
25	26.6	20.2	11561	1	US-08-450-332-1
26	26.6	20.2	11561	2	US-08-637-640-1
27	26.6	20.2	11561	4	US-09-004-406C-1

28	26.4	20.0	28804	2	US-08-592-874-1	Sequence 1, Appl
29	26.4	20.0	28804	3	US-09-096-942-2	Sequence 2, Appl
30	26.4	20.0	28804	3	US-09-096-867-2	Sequence 2, Appl
31	26	19.7	726	3	US-08-660-645A-11	Sequence 11, Appl
32	26	19.7	726	3	US-09-298-718-11	Sequence 11, Appl
33	26	19.7	726	4	US-09-546-969-11	Sequence 11, Appl
34	26	19.7	726	4	US-08-980-832-28	Sequence 28, Appl
35	26	19.7	1113	4	US-09-172-353-1	Sequence 1, Appl
36	26	19.7	1536	4	US-09-352-990-17	Sequence 17, Appl
37	26	19.7	2632	2	US-08-899-324-13	Sequence 32, Appl
38	26	19.7	2632	2	US-08-329-892B-32	Sequence 32, Appl
39	26	19.7	3023	4	US-09-593-589-10	Sequence 10, Appl
40	25.8	19.5	254	4	US-09-328-111-610	Sequence 610, App
41	25.8	19.5	1995	4	US-08-425-069-3	Sequence 3, Appl
42	25.8	19.5	1995	2	US-08-317-844B-3	Sequence 11, Appl
43	25.6	19.4	1103	3	US-08-617-860B-5	Sequence 5, Appl
44	25.6	19.4	2943	1	US-08-042-747A-7	Sequence 7, Appl
45	25.4	19.2	710	4	US-08-998-416-817	Sequence 817, App

ALIGNMENTS

RESULT 1
US-09-328-111-529/c
Sequence 529, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell III, Eddie
APPLICANT: Carroll, Theodore J.
APPLICANT: Dertl, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT FILING DATE: 1999-06-08
CURRENT APPLICATION NUMBER: US/09/328,111
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 529
LENGTH: 251
TYPE: DNA
ORGANISM: Homo sapiens
US-09-328-111-529

Query Match 24.2%; Score 32; DB 4; Length 251;
Best Local Similarity 55.4%; Pred. No. 0.11;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtctgacgagatttcccggttgatctggatctgacatcgtcgttcgcg 64
DB 132 TAGAAGCCTGACATCTCTGCGCCGACATCTAGTCTCATTCGCCCTCCATCC 73
QY 65 tggcgttgatctgacacagcagatcccccaggggtgacatcgtcgttcgcg 116
DB 72 TAGCATTCTTACATATACAGAGGATCAATCCCTCCCTTACCATCA 21

RESULT 2
US-09-385-982-410
Sequence 410, Application US/09385982
Patent No. 6262334

```

; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 344
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 410
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-410
```

```

Query Match          24.2%; Score 32; DB 4; Length 353;
Best Local Similarity 55.4%; Pred. No. 0.12;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

```

QY 5 tggaaatcgtctgagcagatgttcgcggttgatctgtgtctcactcagtgcttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 tagaaacgcgtctgacatctcgtccgcacatctcctagctccatccatccatccc 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 tgcgcgtgctgttcaaacagcagagttcccccagaggtgacatcgatcaca 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 tacgcattcttataaacagcagagtgacatccctccctaccacaa 330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 3

```

US-08-219-842-2
; Sequence 2, Application US/08219842
; Patent No. 5565323
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; TITLE OF INVENTION: for Alzheimer's Disease
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/219,842
; FILING DATE: 30-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-AG 9504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```

; TOPOLOGY: linear
US-08-219-842-2
```

```

Query Match          24.2%; Score 32; DB 1; Length 754;
Best Local Similarity 55.4%; Pred. No. 0.16;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

```

QY 5 tggaaatcgtctgagcagatgttcgcggttgatctgtgtctcactcagtgcttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 TAGAACCCTGTGACTATCTCTGCCGCATCATCTTACTTCTCATCCCTCCATCCC 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 tgcgcgtgctgttcaaacagcagagttcccccagaggtgacatcgatcaca 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 TACGCATCTTTACATATAACAGCAGAGTCAACGATCTCCCTTACATCA 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 4

```

US-08-451-096-2
; Sequence 2, Application US/08451096
; Patent No. 5760205
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; TITLE OF INVENTION: for Alzheimer's Disease
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,096
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/219,842
; FILING DATE: 30-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-AG 9504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 754 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-451-096-2
```

```

Query Match          24.2%; Score 32; DB 1; Length 754;
Best Local Similarity 55.4%; Pred. No. 0.16;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

```

QY 5 tggaaatcgtctgagcagatgttcgcggttgatctgtgtctcactcagtgcttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 TAGAACCCTGTGACTATCTCTGCCGCATCATCTTACTTCTCATCCCTCCATCCC 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 tgcgcgtgctgttcaaacagcagagttcccccagaggtgacatcgatcaca 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 TACGCATCTTTACATATAACAGCAGAGTCAACGATCTCCCTTACATCA 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 5
US-09-385-982-539/C
Sequence 539, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENDEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS: II
FILE REFERENCE: CCDA-260XX
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 539
LENGTH: 789
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(789)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-539

Query Match 24.2%; Score 32; DB 4; Length 789;
Best Local Similarity 55.4%; Pred. No. 0.16;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 5 tggaaatcgtctgacgaattgttcggtgtgattcgtgttcattcattcgttcgcgc 64
DB 132 TAGAAGCCGTCTGAACATCTGCGCCGACATCTAGTCTCTATCGCCCTCCATCCC 73
QY 65 tggcggtgtgtctcaacagcaagatcccgaggggtgacatcgtcatcaa 116
DB 72 TAGCGATCCTTACATACAGACGAGGTCAAGATCCTCCCTTACCATCAA 21

RESULT 6
US-08-413-740A-2
Sequence 2, Application US/08413740A
Patent No. 6171859
GENERAL INFORMATION:
APPLICANT: HERRNSTADT, CORINNA
APPLICANT: PARKER, WILLIAM D.
APPLICANT: DAVIS, ROBERT
APPLICANT: MILLER, SCOTT W.
TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
TITLE OF INVENTION: Defects
NUMBER OF SEQUENCES: 206
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036-5405
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413,740A
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04063
FILING DATE: 30-MAR-1995
APPLICATION NUMBER: 08/413,740
FILING DATE: 30-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bonham, David B.
REGISTRATION NUMBER: 34297
REFERENCE/DOCKET NUMBER: 2105/7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 429-1776
TELEFAX: (202) 429-0796
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 854 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-413-740A-2

Query Match 24.2%; Score 32; DB 4; Length 854;
Best Local Similarity 55.4%; Pred. No. 0.16;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 5 tggaaatcgtctgacgaattgttcggtgtgattcgtgttcattcattcgttcgcgc 64
DB 252 TAGAAGCCGTCTGAACATCTGCGCCGACATCTAGTCTCTATCGCCCTCCATCCC 311
QY 65 tggcggtgtgtctcaacagcaagatcccgaggggtgacatcgtcatcaa 116
DB 312 TAGCGATCCTTACATACAGACGAGGTCAAGATCCTCCCTTACCATCAA 363

RESULT 7
PCT-US95-04063-2
Sequence 2, Application PC/TUS9504063
GENERAL INFORMATION:
APPLICANT: HERRNSTADT, CORINNA
APPLICANT: PARKER, WILLIAM D.
APPLICANT: DAVIS, ROBERT
APPLICANT: MILLER, SCOTT W.
TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
TITLE OF INVENTION: Defects
NUMBER OF SEQUENCES: 206
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036-5405
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04063
FILING DATE: 30-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bonham, David B.
REGISTRATION NUMBER: 34297
REFERENCE/DOCKET NUMBER: 2105/7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 429-1776

TELEFAX: (202) 429-0796
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 854 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 PCT-US95-04063-2

Query Match
 Best Local Similarity 24.2%; Score 32; DB 5; Length 854;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtcgaacgattgtccggtgtgattcgtgtcattcgtgtcgtcgc 64
 DB 252 TAGAACCCTGTGAACACTCTCTGCGCCGATCATCTAGTCTCATGCGCCCTCCATCCC 311
 QY 65 tgcggtgctgttcaaacagcaagagttcccgagggtgacatcgtcatcaa 116
 DB 312 TACGCATCCTTTACATTAACAGACGAGTCAACGATCCCTCCATTACCATCAA 363

RESULT 8

US-09-097-889-2
 Sequence 2, Application US/09097889
 Patent No. 6218117

GENERAL INFORMATION:
 APPLICANT: Herrstadt, Corrina
 APPLICANT: Ghosh, Soumitra S.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
 TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
 NUMBER OF SEQUENCES: 26
 CURRENT APPLICATION DATA: EXTRA MITOCHONDRIAL DNA RATIOS
 FILING DATE: 15-JUN-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rosenman Ph.D., Stephen J.
 REGISTRATION NUMBER: 43,058
 REFERENCE/DOCKET NUMBER: 660088.417
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 16569 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-097-889-2

Query Match
 Best Local Similarity 24.2%; Score 32; DB 4; Length 16569;
 Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtcgaacgattgtccggtgtgattcgtgtcattcgtgtcgtcgc 64
 DB 7767 TAGAACCCTGTGAACACTCTCTGCGCCGATCATCTAGTCTCATGCGCCCTCCATCCC 7826
 QY 65 tgcggtgctgttcaaacagcaagagttcccgagggtgacatcgtcatcaa 116
 DB 7827 TACGCATCCTTTACATTAACAGACGAGTCAACGATCCCTCCATTACCATCAA 7878

RESULT 9

US-09-097-889-1
 Sequence 1, Application US/09097889
 Patent No. 6218117

GENERAL INFORMATION:
 APPLICANT: Herrstadt, Corrina
 APPLICANT: Ghosh, Soumitra S.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
 TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
 NUMBER OF SEQUENCES: 26
 CURRENT APPLICATION DATA: EXTRA MITOCHONDRIAL DNA RATIOS
 FILING DATE: 15-JUN-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rosenman Ph.D., Stephen J.
 REGISTRATION NUMBER: 43,058
 REFERENCE/DOCKET NUMBER: 660088.417
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6744 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-097-889-1

Query Match
 Best Local Similarity 23.6%; Score 31.2; DB 4; Length 6744;
 Matches 60; Conservative 2; Mismatches 50; Indels 0; Gaps 0;

QY 5 tggaaatcgtcgaacgattgtccggtgtgattcgtgtcattcgtgtcgtcgc 64
 DB 4343 TAGAACCCTGTGAACACTCTCTGCGCCGATCATCTAGTCTCATGCGCCCTCCATCCC 4402
 QY 65 tgcggtgctgttcaaacagcaagagttcccgagggtgacatcgtcatcaa 116
 DB 4403 TACGCATCCTTTACATTAACAGACGAGTCAACGATCCCTCCATTACCATCAA 4454

RESULT 10

US-09-156-979-1
 Sequence 1, Application US/09156979
 Patent No. 5962672

GENERAL INFORMATION:
 APPLICANT: Cowser, Lex M.


```

;          TOPOLOGY: linear
;          MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

```

62 cgctgccggtgtcgttcaacacgacagagttccccgaggtgacatcgttcatcaac

270
 280
 290
 300
 310
 320
 330
 340
 350
 360
 370
 380
 390
 400
 410
 420
 430
 440
 450
 460
 470
 480
 490
 500
 510
 520
 530
 540
 550
 560
 570
 580
 590
 600
 610
 620
 630
 640
 650
 660
 670
 680
 690
 700
 710
 720
 730
 740
 750
 760
 770
 780
 790
 800
 810
 820
 830
 840
 850
 860
 870
 880
 890
 900
 910
 920
 930
 940
 950
 960
 970
 980
 990
 1000
 1010
 1020
 1030
 1040
 1050
 1060
 1070
 1080
 1090
 1100
 1110
 1120
 1130
 1140
 1150
 1160
 1170
 1180
 1190
 1200
 1210
 1220
 1230
 1240
 1250
 1260
 1270
 1280
 1290
 1300
 1310
 1320
 1330
 1340
 1350
 1360
 1370
 1380
 1390
 1400
 1410
 1420
 1430
 1440
 1450
 1460
 1470
 1480
 1490
 1500
 1510
 1520
 1530
 1540
 1550
 1560
 1570
 1580
 1590
 1600
 1610
 1620
 1630
 1640
 1650
 1660
 1670
 1680
 1690
 1700
 1710
 1720
 1730
 1740
 1750
 1760
 1770
 1780
 1790
 1800
 1810
 1820
 1830
 1840
 1850
 1860
 1870
 1880
 1890
 1900
 1910
 1920
 1930
 1940
 1950
 1960
 1970
 1980
 1990
 2000
 2010
 2020
 2030
 2040
 2050
 2060
 2070
 2080
 2090
 2100
 2110
 2120
 2130
 2140
 2150
 2160
 2170
 2180
 2190
 2200
 2210
 2220
 2230
 2240
 2250
 2260
 2270
 2280
 2290
 2300
 2310
 2320
 2330
 2340
 2350
 2360
 2370
 2380
 2390
 2400
 2410
 2420
 2430
 2440
 2450
 2460
 2470
 2480
 2490
 2500
 2510
 2520
 2530
 2540
 2550
 2560
 2570
 2580
 2590
 2600
 2610
 2620
 2630
 2640
 2650
 2660
 2670
 2680
 2690
 2700
 2710
 2720
 2730
 2740
 2750
 2760
 2770
 2780
 2790
 2800
 2810
 2820
 2830
 2840
 2850
 2860
 2870
 2880
 2890
 2900
 2910
 2920
 2930
 2940
 2950
 2960
 2970
 2980
 2990
 3000
 3010
 3020
 3030
 3040
 3050
 3060
 3070
 3080
 3090
 3100
 3110
 3120
 3130
 3140
 3150
 3160
 3170
 3180
 3190
 3200
 3210
 3220
 3230
 3240
 3250
 3260
 3270
 3280
 3290
 3300
 3310
 3320
 3330
 3340
 3350
 3360
 3370
 3380
 3390
 3400
 3410
 3420
 3430
 3440
 3450
 3460
 3470
 3480
 3490
 3500
 3510
 3520
 3530
 3540
 3550
 3560
 3570
 3580
 3590
 3600
 3610
 3620
 3630
 3640
 3650
 3660
 3670
 3680
 3690
 3700
 3710
 3720
 3730
 3740
 3750
 3760
 3770
 3780
 3790
 3800
 3810
 3820
 3830
 3840
 3850
 3860
 3870
 3880
 3890
 3900
 3910
 3920
 3930
 3940
 3950
 3960
 3970
 3980
 3990
 4000
 4010
 4020
 4030
 4040
 4050
 4060
 4070
 4080
 4090
 4100
 4110
 4120
 4130
 4140
 4150
 4160
 4170
 4180
 4190
 4200
 4210
 4220
 4230
 4240
 4250
 4260
 4270
 4280
 4290
 4300
 4310
 4320
 4330
 4340
 4350
 4360
 4370
 4380
 4390
 4400
 4410
 4420
 4430
 4440
 4450
 4460
 4470
 4480
 4490
 4500
 4510
 4520
 4530
 4540
 4550
 4560
 4570
 4580
 4590
 4600
 4610
 4620
 4630
 4640
 4650
 4660
 4670
 4680
 4690
 4700
 4710
 4720
 4730
 4740
 4750
 4760
 4770
 4780
 4790
 4800
 4810
 4820
 4830
 4840
 4850
 4860
 4870
 4880
 4890

00-800-644-93

[illegible]

```

RESULT 15
US-08-776-210-3
; Sequence 3, Application US/08776210
; Patent No. 5942659
; GENERAL INFORMATION:
; APPLICANT: ALIBERT, Gilbert
; APPLICANT: MOULOUNGUI, Zephirin
; APPLICANT: BOUDET, Alain
; TITLE OF INVENTION: PROCESS FOR PRODUCING FATTY ACIDS OR
; DERIVATIVES THEREOF FROM OLEAGINOUS PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,210
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94 09272
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR95/00957
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: IN 387 - BE 6996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-776-210-3

```

```

Query Match 20.9% Score 27.6; DB 2; Length 1600;
Best Local Similarity 63.6%; Pred. No. 5.3;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 57 gtctcgctgcgcggtctcaacagcaagagttcccgaggggtgacatcatcaaa 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 894 gtctcgctgcgcggtctcaacagcaagagttcccgaggggtgacatcatcaaa 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 117 cgtcaga 122
   | |||
DB 954 CAACGA 959

```

Search completed: December 22, 2001, 09:36:27
Job time: 5198 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 08:48:25 ; Search time 2235.47 Seconds
(without alignments)
634.517 Million cell updates/sec

Title: US-09-712-768-3
Perfect score: 132
Sequence: 1 ccgcgtggaatcgtctgtgac.....tcaacgtcgtggtctgtac 132

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	36.2	27.4	582	11	BI180715 LRRH22B7 L
2	36.2	27.4	643	11	BI180826 3122B7 LR
3	35.2	26.7	546	10	AM270568 x682h07.x
4	35.2	26.7	640	10	AV702099 AV702099
5	34.4	26.1	244	11	W03206
6	34.4	26.1	283	10	AA176518 2P37D03.r
7	34.4	25.8	588	11	BG901209 950001C09
8	33.6	25.5	162	10	BE175243 OY2-HT057
9	33.6	25.5	177	10	BE180750 RC3-HT062
10	33.6	25.5	191	11	BG876169 MR1-CT035
11	33.6	25.5	198	11	BG876149 MR1-CT035
12	33.6	25.5	200	11	BG876153 MR1-CT035

13	33.6	25.5	202	11	BG876161	BG876161 MR1-CT035
14	33.6	25.5	205	11	BG876168	BG876168 MR1-CT035
15	33.6	25.5	258	11	BG875536	BG875536 RC6-CN000
16	33.6	25.5	264	11	BG877167	BG877167 QV3-HR045
17	33.6	25.5	295	11	C17077	C17077 C17077 Clon
18	33.6	25.5	303	10	AA879300	AA879300 nwb6f08.s
19	33.6	25.5	325	11	BF174777	BF174777 MYE3800 M
20	33.6	25.5	361	11	BF170586	BF170586 PCL1123 M
21	33.6	25.5	403	11	BG876871	BG876871 QV0-HR036
22	33.6	25.5	462	11	BF172424	BF172424 PCL4905 M
23	33.6	25.5	525	11	C17914	C17914 C17914 Huma
24	33.6	25.5	546	11	BG875704	BG875704 CM0-CN004
25	33.6	25.5	554	10	A1028076	A1028076 o43d10.x
26	33.6	25.5	625	10	AV711246	AV711246 AV711246
27	33.6	25.5	662	10	AV727896	AV727896 AV727896
28	33.6	25.5	668	10	AV713349	AV713349 AV713349
29	33.6	25.5	731	10	AV762348	AV762348 AV762348
30	33.6	25.5	836	2	BG529384	BG529384 602558517
31	33.4	25.3	536	11	BE971012	BE971012 601651228
32	33.4	25.3	691	10	AV701188	AV701188 AV701188
33	33.4	25.3	716	10	A1525178	A1525178 promrna-9
34	33.4	25.3	861	11	BE878356	BE878356 601488093
35	33.2	25.2	187	10	AA092360	AA092360 117586.se
36	33.2	25.2	246	11	B1032317	B1032317 CM3-NN025
37	33.2	25.2	339	10	AA179137	AA179137 zpa1f12.r
38	33.2	25.2	517	10	AA196841	AA196841 z909903.s
39	33.2	25.2	776	11	BG116705	BG116705 602317417
40	33.2	25.0	277	10	BE044951	BE044951 hn08h03.x
41	33.2	25.0	738	11	BG612163	BG612163 602614124
42	32.6	24.7	854	13	CNS04005	ALJ02918 Tetracodon
43	32.6	24.7	198	11	BG876186	BG876186 MR1-CT035
44	32.6	24.7	323	10	A1160825	A1160825 qc77a04.x
45	32.6	24.7	364	10	AA229353	AA229353 nc47e03.r

ALIGNMENTS

RESULT 1
LOCUS BI180715 582 bp mRNA EST 09-JUL-2001
DEFINITION LRRH22B7 LRH (Lin Rhodamine and Hoechst dye) Mus musculus cDNA clone
ACCESSION 3217089(5'), mRNA sequence.
VERSION BI180715
KEYWORDS BI180715.1 GI:14646526
SOURCE EST.
ORGANISM house mouse.
MUS musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 582)
AUTHORS Ma,X.Y., Husain,T., Peng,H., Lin,S., Lu,B.F., Tuck,D., Mironenko,O., Johnson,S., Krause,D. and Perkins,A.
TITLE Development of a murine myeloid cDNA microarray using a subtracted cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Hui Peng, MD, & Phd.
Dr. Archibald Perkins' lab
Yale University School of Medicine
310 Cedar Street, LH 305, New Haven, CT 06520, USA
Tel: 203-764-9977
Email: hui.peng@yale.edu
This clone is available royalty-free through LNL: contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13AK as 3' primer, T7 as 5' primer.
Location/Qualifiers
1. 582
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="3217089(5)"
/clone_lib="LRH (Lin Rhodamine and Hoechst dye)"
/sex="Female"

```

/tissue_type="Primary sorted bone marrow cells"
/dev_stage="Adult"
/lab_host="DH10B (phage-resistant)"
/Note="Vector: pZLI, Site_1: SalI site; Site_2: EagI site; LRH library was constructed from cDNA of primary bone marrow cells depleted of lineage-committed cells and enriched for primitive cells by FACS sorting for cells with low level staining with rhodamine123 and Hoechst 33342 dyes. cDNA from 5000 cells derived from 30 mice, 5000 cells were directionally cloned into SalI-EagI restriction site of the ( Ziplox (Gibco BRL). The ligated cDNA fragments were transformed into DH10B host cells. The original library had an initial plating complexity of 1.44x107 clones."
BASE COUNT      136 a      164 c      161 g      119 t      2 others
ORIGIN

Query Match      27.4%; Score 36.2; DB 11; Length 582;
Best Local Similarity 56.2%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 5 tgaatcgtctggaacgtatgttcggtgtgtatctgtcttcacgtgtcgttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 339 TGGAAATCCTGTGACCATCGCCCTTGTGATCTGTGTAATGCGATTCGCCGA 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 tgcggtgtctgtcaaacgacagatgcccgagggtgacatgtcatcaacgtcgag 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 399 CCGCCACGCTGATCAAGATGTACGACACGATGCGACGTGATTCAGATCAGT 458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 125 g 125
DB 459 G 459

RESULT 2
BI180826 643 bp mRNA EST 09-JUL-2001
LOCUS 312287 LRH (lin Rhodamine and Hoechst dye) Mus musculus cDNA clone
DEFINITION 3217089(3'), mRNA sequence.
ACCESSION BI180826
VERSION BI180826.1 GI:14646637
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 643)
Ma,X.Y., Husain,T., Peng,H., Lin,S., Lu,B.F., Tuck,D., Mironenko,O.,
Johnson,S., Krause,D. and Perkins,A.
Development of a murine myeloid cDNA microarray using a subcloned
cDNA library
Unpublished (2001)
Contact: Hui Peng, MD, 6 PhD.
Dr. Archibald Perkins' Lab
Yale University School of Medicine
310 Cedar Street, LH 305, New Haven, CT 06520, USA
Tel: 203-764-9977
Email: hui.peng@yale.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: MJ3ACK as 3' primer, T7 as 5' primer.
Location/Qualifiers
1..643
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="3217089(3)"
/clone_lib="LRH (Lin Rhodamine and Hoechst dye)"
/sex="Female"
/tissue_type="Primary sorted bone marrow cells"
/dev_stage="Adult"
/lab_host="DH10B (phage-resistant)"

```

```

/Note="Vector: pZLI, Site_1: SalI site; Site_2: EagI site; LRH library was constructed from cDNA of primary bone marrow cells depleted of lineage-committed cells and enriched for primitive cells by FACS sorting for cells with low level staining with rhodamine123 and Hoechst 33342 dyes. cDNA from 5000 cells derived from 30 mice, 5000 cells were directionally cloned into SalI-EagI restriction site of the ( Ziplox (Gibco BRL). The ligated cDNA fragments were transformed into DH10B host cells. The original library had an initial plating complexity of 1.44x107 clones."
BASE COUNT      134 a      179 c      171 g      153 t      6 others
ORIGIN

Query Match      27.4%; Score 36.2; DB 11; Length 643;
Best Local Similarity 56.2%; Pred. No. 1.7;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 5 tgaatcgtctggaacgtatgttcggtgtgtatctgtcttcacgtgtcgttcgcg 64
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 269 TGGAAATCCTGTGACCATCGCCCTTGTGATCTGTGTAATGCGATTCGCCGA 210
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 tgcggtgtctgtcaaacgacagatgcccgagggtgacatgtcatcaacgtcgag 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 209 CCGCCACGCTGATCAAGATGTACGACACGATGCGACGTGATTCAGATCAGT 150
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 125 g 125
DB 149 G 149

RESULT 3
AM270568 546 bp mRNA EST 03-JAN-2000
LOCUS xp82h07.x1 NCI-CGAP_Ov40 Homo sapiens cDNA clone IMAGE:2746909 3'
DEFINITION similar to SW:COX2_HUMAN P00403 CYTOCHROME C OXIDASE POLYPEPTIDE II
; mRNA sequence.
ACCESSION AM270568
VERSION AM270568.1 GI:6657598
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 546)
NCI-CGAP http://www.ncbi.nlm.nih.gov/cgi/gap
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Andrew Berchuck M.D., John Gillespie M.D.,
Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html
Possible reversed clone: similarity on wrong strand
Seq primer: -400P from Gibco
High quality sequence stop: 397.
Location/Qualifiers
1..546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2746909"
/clone_lib="NCI-CGAP_Ov40"
/sex="Female"
/tissue_type="endometrialoid ovarian metastasis"
/lab_host="DH10B"

```

/note="Organ: ovary; Vector: PAM10; cDNA made by oligo-dt priming. Non-directionally cloned into the UDG sites of PAM10. Size-selected on agarose gel, average insert size 500 bp. Primary library, non-amplified. cDNA Library Preparation: David B. Krizman, Ph.D (NCI). Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 151 a 172 c 90 g 133 t
ORIGIN

Query Match 26.7%; Score 35.2; DB 10; Length 546;
Best Local Similarity 57.1%; Pred. No. 3.2;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

OY 5 tggaaatcgtcgtgagcaggtgttcgcgtgtgattcgttcattcagtgagtcgtcgc 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 179 TAGAACCCTCTGACATATCTCTGCCCGCATCTATGATCTCATCGCCCTCTATCCC 238

OY 65 tgcgcgtcgtctcaaacagcagagttcccgaggggtgacatcgtcatca 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TACGATCTTTTACATACAGACGAGTCACAGACCCCTCCCTTACATCA 290

RESULT 4
AV702099 640 bp mRNA EST 08-OCT-2000
LOCUS
DEFINITION AV702099 ADB Homo sapiens cDNA clone ADBCV12 5', mRNA sequence.
ACCESSION AV702099
VERSION AV702099.1 GI:10718429
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 640)
AUTHORS Peng,Y., Song,H., Huang,Q., Huang,C., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,Y., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R., Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA ADB clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
331 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@cnhg.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. 640
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ADBCV12"
/clone_lib="ADB"
/tissue_type="Adrenal gland"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 194 a 195 c 95 g 154 t 2 others
ORIGIN

Query Match 26.7%; Score 35.2; DB 10; Length 640;
Best Local Similarity 57.1%; Pred. No. 3.3;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

OY 5 tggaaatcgtcgtgagcagatgttcgcgtgtgattcgttcattcagtgagtcgtcgc 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71 TAGAACCCTCTGACATATCTCTGCCCGCATCTATGATCTCATCGCCCTCTATCCC 130

OY 65 tgcgcgtcgtctcaaacagcagagttcccgaggggtgacatcgtcatca 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 TACGATCTTTTACATACAGACGAGTCACAGACCCCTCCCTTACATCA 182

RESULT 5
W03206 244 bp mRNA EST 18-APR-1996
LOCUS
DEFINITION z64b03.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:297293 5' similar to SW:COX2.HUMAN P00403 CYTOCHROME C
OXIDASE POLYPEPTIDE II, mRNA sequence.
W03206
VERSION W03206.1 GI:1275224
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 244)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the IMAG Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone; similarity on wrong strand
Seq primer: mob.REGA+ET
High quality sequence stop: 1.
Location/Qualifiers
1. 244
/organism="Homo sapiens"
/db_xref="GDB:124214"
/db_xref="taxon:9606"
/clone="IMAGE:297293"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: liver and spleen; Vector: pF7T3D (Pharmacia) with a modified polylinker; site_1: Pac I; site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer (5' ACTGGAAGATTTATTTAATTAAGATCTTTTATTTTATTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pF7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonalao."

BASE COUNT 65 a 64 c 27 g 61 t 27 others
ORIGIN

Query Match 26.1%; Score 34.4; DB 11; Length 244;
Best Local Similarity 52.7%; Pred. No. 4.7;
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

OY 5 tggaaatcgtcgtgagcagatgttcgcgtgtgattcgttcattcagtgagtcgtcgc 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 TAGAACCCTCTGACATATCTCTGCCCGCATCTATGATCTCATCGCCCTCTATCCC 167

OY 65 tgcgcgtcgtctcaaacagcagagttcccgaggggtgacatcgtcatca 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 TACGATCTTTTACATACAGACGAGTCACAGATNNNNNNNTAGATCA 219

MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV2-HT0577-08) 500-211-c09&t3=2000-05-08&f4=1) Seq primer: puc 18 forward High quality sequence start: 11 High quality sequence stop: 162. Location/Qualifiers 1..162 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HT0577" /dev_stage="Adult" /node="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."BASE COUNT 52 a 59 c 17 g 34 t ORIGIN
Query Match	25.5%; Score 33.6; DB 10; Length 162;
Best Local Similarity	56.2%; Pred. No. 7.4;
Matches	63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
OY	5 tggaaatcgtctgtagagattgtccggttgatctcgcttcacgtgcgtctcgc 64 Db 48 TAGAAGACGCTTGAACTATCGTCCGCCGCATCATCTCCTAGTCCTCAWCCGCCCTCCCATCCC 107
OY	65 tgcgcgtcgtctccaacagcaagaagttcccgcaggaggtgacatcgtaactcaa 116 Db 108 TACGATCTCTTTACTATTACAGACAGACGATGAAGATCCCTCCCTCACATCAA 159
RESULT 9	
LOCUS	BE180750 177 bp mRNA EST 22-JUN-2000
DEFINITION	RC3-HIT0626-130400-022-f12 HT0626 Homo sapiens cDNA, mRNA sequence.
VERSION	BE180750
KEYWORDS	BE180750.1 GI:8659926
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 177)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagal,M.A., da Silva,W Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

FEATURES	source	Location/Qualifiers
BASE COUNT	53 a	62 c 19 g 43 t
ORIGIN		
Query Match	25.5%	Score 33.6; DB 10; Length 177;
Best Local Similarity	56.2%	Pred. No. 7.5;
Matches	63; Conservative	0; Mismatches 49; Indels 0; Gaps 0;
OY	5	tggaaatcgtctgtagagatgtgtccggttgatctggtcttcacgtgtagtctgcg 64
Db	64	TAGAAATGTCGTGAACACTATCCGCGCCGCATCATCTCTATCCGCCATCCATCC 123
OY	65	tggcgggtcgtcgtctgaacagcaagaaggtcccgagggtagaatgcatatcaa 116
Db	124	TACGCATCTCTTTACTATACAGACAGAGTCAACAGATCCCTCCCTTACATCA 175
RESULT	10	
LOCUS	BG876169	191 bp mRNA EST 30-MAY-2001
DEFINITION	MR1-CT0352-230200-101-e09 CT0352 Homo sapiens cDNA, mRNA sequence.	
ACCESSION	BG876169	
VERSION	BG876169.1	GI:14253259
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates; Catarrhini; Homnidae: Homo.	
AUTHORS	1 (bases 1 to 191) Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldmann,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunsstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.	
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	
MEDLINE	20202663	
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=RC3-HT0626-130)	

200-101-e09&t3=2000-02-23&t4=1)	
Seq primer: puc 18 forward	
High quality sequence start: 2	
High quality sequence stop: 191	
Location/Qualifiers	

/note="Organ: colon; Vector: puc18; Site:1: Smal; site:2: Sma1; A mini-library was made by cloning products derived from ORESRES PCR (U.S. Letters Patent application No. 166,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

Oy 65 tgcgggtgctgttcaaacagcagaagtcccccgagggtgatcatgctatcaa 116
| | | | | | | | | | | |
Db 101 TACGCATCCCTTACATATACAGACGAGCAGTCACGATGCCCTCCTTACCATCAA 152

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0352"
/dev_stage="Adult"
/name="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESFES PCR (U.S. letters Patent application No. 1967,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      58 a      68 c      29 g      43 t
ORIGIN

```

	Query Match	Score	DB	Length
Best Local Similarity	25.5%	33.6	11	198
Matches	63	Conservative	0	Mismatches 49; Indels 0; Gaps 0;
OY	5	ttggaatcgtctcgagcagatgttcgcggttgatctatgctctcaatcgatgctctcgc	64	
Db	48	tgcgaaccccttgcagactatctcgtccggccgcatcatctatctccatgcgccctccatccc	107	
OY	65	tgcgcgtctcgttcaaacgcagcaagttcccccaggagtgacacgtcatcaaa	116	
Db	108	tacgcattcctttacattacacgacagagcttaaacgattccctcttaccatcaa	159	

RESULT	12				
LOCUS	BG876153				
DEFINITION	BG876153	200 bp	mRNA	EST	30-MAY-2001
ACCESSION	BM1-C70352	-220200-102-c07	C70352	Homo sapiens	cdNA, mRNA sequence.
VERSION	BG876153				
KEYWORDS	BG876153.1	GI:14253243			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
	1 (bases 1 to 200)				
	Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M.R., Nagai, M.A., de Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsushima, A., Bala, G.S., Simpson, D.H., Brundin, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)	
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G.				

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CP0352"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; site_1: SmaI; site_2:
SmaI; A mini-library was made by cloning products derived

```

/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 59 a 68 g 28 g 45 t

ORIGIN

Query Match 25.5%; Score 33.6; DB 11; Length 200;
 Best Local Similarity 56.2%; Pred. No. 7.7;
 Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 5 tggaaatcgtctggaagatgttccggtgtgattctgtcttcacgtgcttcgc 64
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 52 TGAACCCGCTGAACATACCTGCGCCGATCATCTAGTCCATGCCCTCCATCCC 111

Qy 65 tggcgtgtctgttcaacgaagaagttccggagggtgacatcgtcatcaa 116
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 110 TACGCATCTTACATATACAGACGAGGTCAACGATCCCTCCATCCATCAA 161

RESULT 13

LOCUS BG876161 202 bp mRNA EST 30-MAY-2001
 DEFINITION MRL-CT0352-220200-102-h12 CT0352 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG876161
 VERSION BG876161.1 GI:14253251
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 202)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MRL-CT0352-220
 200-102-h12&t3=2000-02-22&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 14
 High quality sequence stop: 202.
 Location/Qualifiers

FEATURES

source 1..202
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="CT0352"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2:
 Sma1; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 61 a 69 g 29 g 42 t 1 others

ORIGIN

Query Match 25.5%; Score 33.6; DB 11; Length 202;
 Best Local Similarity 56.2%; Pred. No. 7.7;
 Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 5 tggaaatcgtctggaagatgttccggtgtgattctgtcttcacgtgcttcgc 64
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 52 TGAACCCGCTGAACATACCTGCGCCGATCATCTAGTCCATGCCCTCCATCCC 111

Qy 65 tggcgtgtctgttcaacgaagaagttccggagggtgacatcgtcatcaa 116
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 112 TACGCATCTTACATATACAGACGAGGTCAACGATCCCTCCATCCATCAA 163

RESULT 14

LOCUS BG876168/c 205 bp mRNA EST 30-MAY-2001
 DEFINITION MRL-CT0352-230200-101-d11 CT0352 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG876168
 VERSION BG876168.1 GI:14253258
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 205)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MRL-CT0352-230
 200-101-d11&t3=2000-02-23&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 11
 High quality sequence stop: 205.
 Location/Qualifiers

FEATURES

source 1..205
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="CT0352"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2:
 Sma1; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 43 a 29 g 72 g 60 t 1 others

Query Match 25.5%; Score 33.6; DB 11; Length 205;
 Best Local Similarity 56.2%; Pred. No. 7.7;
 Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 5 tggaaatcgtctggaacgatttcccggtgtgattcgtgtcttcacgtggttcgc 64
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 165 TGGAAACCGTCTGAACATCTCCCGCCATCATCTCTCATCTCCCTCCCATCC 106
 QY 65 tggcggctctgtctcaacgacgaagttcccgagagggtgacatcgtcatca 116
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 105 TACGCATCTTTACATACAGACGAGTCAACGATCCCTCCCTACATCAA 54

RESULT 15

BG875536/c

LOCUS BG875536 258 bp mRNA EST 30-MAY-2001
 DEFINITION RC6-CN0009-020200-011-F12 CN0009 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG875536

VERSION BG875536.1 GI:14252520

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 258)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=ct2-RC6-CN0009-020>)

200-011-F12ct3-2000-02-02ct4-1)

Seq primer: puc 18 forward

High quality sequence stop: 258.

Location/Qualifiers

1. 258

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CN0009"

/dev_stage="Adult"

/note="Organ: colon; normal: Vector: puc18; Site: 1: SmaI;

Site: 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

Db 152 TACGCATCTTTACATACAGACGAGTCAACGATCCCTCCCTACATCAA 101

Search completed: December 22, 2001, 08:48:27
 Job time: 2448 sec

BASE COUNT 55 a 34 c 89 g 80 t

Query Match 25.5%; Score 33.6; DB 11; Length 258;

Best Local Similarity 56.2%; Pred. No. 8;

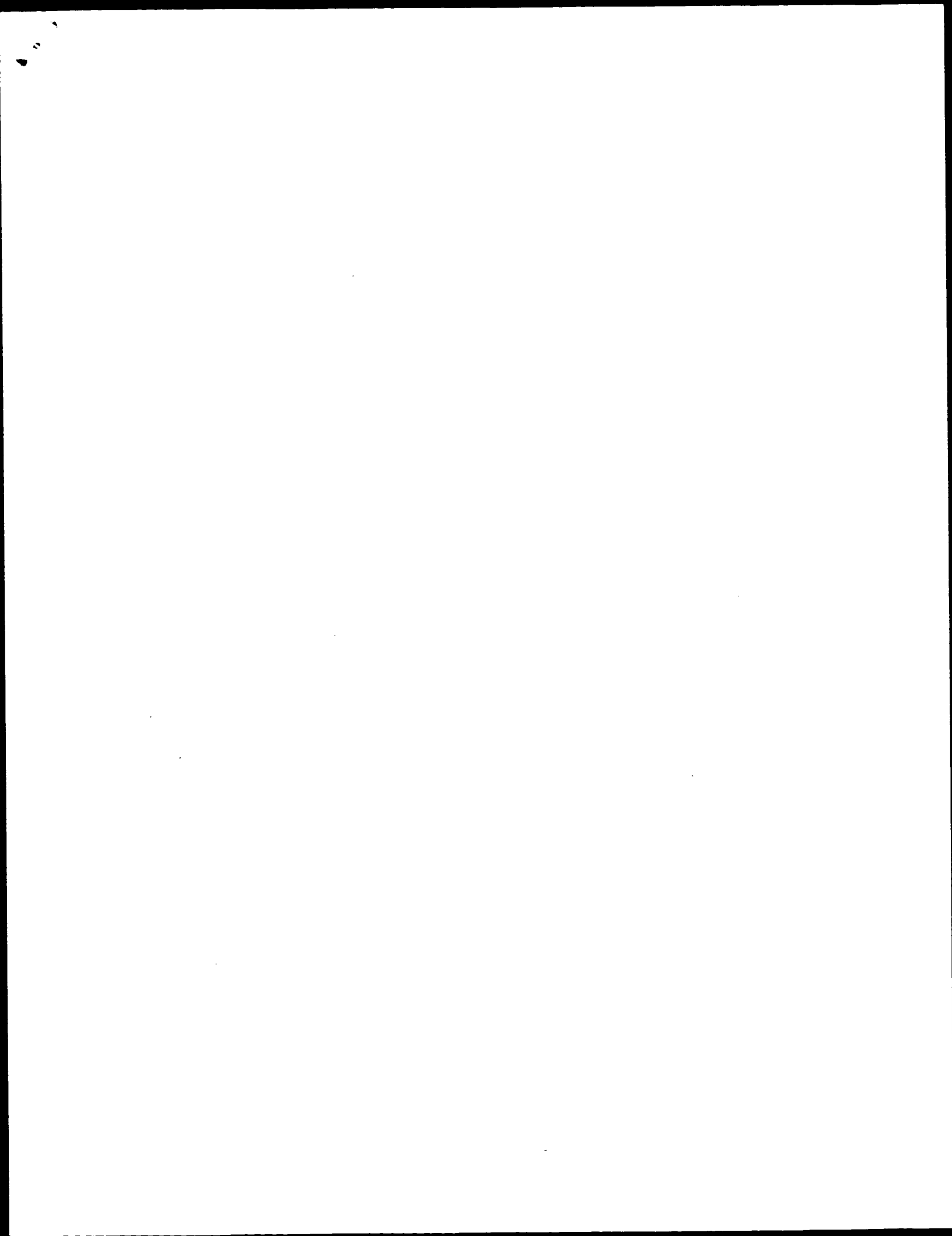
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 5 tggaaatcgtctggaacgatttcccggtgtgattcgtgtcttcacgtggttcgc 64
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 212 TGGAAACCGTCTGAACATCTCCCGCCATCATCTCTCATCTCCCTCCCATCC 153

QY 65 tggcggctctgtctcaacgacgaagttcccgagagggtgacatcgtcatca 116
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Thu Dec 27 08:21:46 2001

us-09-712-768-3.rst




```

|||||
201 AGGCGACGTCGATCGCGGATGCGCTCGACAGACATGTCAGCGCAACGAC 250
84 1sleuTrpAsnValMetValThrTrpHisGlyIleLeuMetCphePhe 100
251 ACCTGTGAACATCATGGTTACTACCACTGATATTCATGATGATGTTCTT 300
101 ValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTrpIleLeuMetProLe 117
301 GTGGGATACCCCGCATTTCTCGGTGGTTTGGTAATCTTCATGATGCCGCT 350
117 uGlnIleGlyAlaProAspMetAlaIlePheProArgMetAsnAsnLeuSerP 134
351 GCAAAATCGCGCTCGGATATGCGCTTCCCGCATGACCAACCTGTCGT 400
134 heTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAla 150
401 TCTGCTGTTGATGCGCGTACCGGATGCGGATGCGCTTCCGCTTCCGCA 450
151 ProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTrpValLeuTrpPr 167
451 CCGGGCGGTGACGGTCACTGGGTTCCGGGCTTGGTTGGTTCTGTACCC 500
167 oProLeuSerThrArgGlnAlaGlyTrpSerMetAspLeuAlaIlePheA 184
501 GCCGCTGTCGACCGCGCAACCTGGCTATTGATGACCTCCGCAATTTTCG 550
184 1aValHisLeuSerGlyValaSerSerIleMetGlyAlaIleAsnMetIle 200
551 CGGTTCACCTGTGCGGTGCTCTCTGATCATAGGCGCGCATCAACATATTC 600
201 ThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisIleValPr 217
601 ACGACCTTCTTGAACATGCGCGCCCGCATGACGGTGACCAAAAGTGGCC 650
217 oLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLeuLeuAla 234
651 GTTGTTCGCGGTGATCTTATCACGGCTTGGCTGATCTCTGCTGGCGC 700
234 euProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPhe 250
701 TGCCGCTTCTGGCTGGTGCATCATCACTGCTGACCGACCGCACTTAC 750
251 GlyThrThrPhePheAsnProAlaGlyGlyAspProIleLeuTrpG1 267
751 GGCACGACCTTCTCAATCTGCTGGCGGCGGTGACCGCATTTCTGACA 800
267 nHisIleLeuTrpPhePheGlyHisProGlyValTrpIleIleLeuLeuP 284
801 ACACATCTCTGTGTTCTTGGGACCGCGAAGTGTACATCATCATTTCTGC 850
284 roGlyPheGlyIleIleSerHisValaSerThrPheSerLysLysPro 300
851 CCGGCTTTGGCATCATGACCATCTGCTGTCGACCTTCTCGAAAACCCG 900
301 ValPheGlyTrpLeuProMetValaTrpAlaMetValaIleGlyValLe 317
901 GTCTTCGCTTACCTGCGGATGCTATGCAATGCTGCAATCGGTGTTCT 950
317 uGlyPheValValTrpAlaHisHisMetTrpThrValGlyMetSerLeuTr 334
951 GGGCTTTGTCTGTGGCGCACACAGTACACCGTTGTGATGTCGCTGA 1000
334 hArgInGlnSerTrpPheMetLeuAlaThrMetValaIleAlaValProThr 350
1001 CCGAGCAATCTCTACTCATGCTGGCCACCAATGATGATCGGTCGCCACC 1050
351 GlyIleLeuLysIlePheSerTrpIleAlaThrMetTrpGlyGlySerValG1 367
1051 GGCATTTAAAGATCTTCTGATGATGCGCACAGATGTTGGGCGGCTGTTGA 1100
367 uPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPheThrV 384
|||||

```

```

1101 GTTCAATCCCGCATGCTCTGGGCTTTGGCTTTATGTTCTGTTACCGG 1150
384 a1GlyGlyValaThrGlyIleValaLeuAlaGlnAlaGlyLeuAspArgAla 400
1151 TGGGTGTTGACCGGATGATCGTCTGCGCCCAAGCGGCTCGACCTGCA 1200
401 TyrHisPheTrpTrpValaValaIleHisPheHisTrpValMetSerLe 417
1201 TATCAGCAGCACTTATTACGTGGTGGCGCACCTTCATTTATGATGTCGT 1250
417 uGlyAlaIlePheAlaIlePheAlaGlyIleTrpPheTrpMetProLysP 434
1251 GGGTGGATGCTTTGCCATGCTTCCCGGTATCATCTTTACATGCCGAAT 1300
434 heSerGlyArgAlaPheProGlyTrpAlaAlaLysLeuHisPheTrpThr 450
1301 TCTGGGCGCGCTTCCCGGAATGGCTGCAAAAGCTGCACATTCTGGACC 1350
451 PhePheIleGlyAlaAsnValaThrPhePheProGlnHisPheLeuGlyAr 467
1351 TTTCTCATGCTGTCGAACGTCACTGTTCTTCCGACGACCTTCTGGGACG 1400
467 gGlnGlyMetProArgTrpArgTrpIleAspTrpProGlnAlaPheAlaLeuT 484
1401 TCAGGGATGCGCGCGCGTTACATGACATGCCGAAGCCTTCCGCGTGT 1450
484 rPasnLysValaSerSerTrpGlyAlaPheLeuAlaPheAlaSerPheLeu 500
1451 GGAACAAAGTCTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
501 PhePheIleValIlePheValaThrThrLeuValaIleAlaArgArgGlyuth 517
1501 TTTCTCATGCTGATCTTGTGATATACGCTGCTGCTGCTGCTGCTGCTG 1550
517 rArgProAsnProTrpGlyGlyPheAlaAspThrLeuGlnTrpThrLeuP 534
1551 CCGTCCGAACCGCTGGGCGCAATTCGCCGATAGCGTGAATGACGACGTGC 1600
534 roSerProProProAlaHisThrPheGlyThrLeuProLysArgSerAsp 550
1601 CATCACCGCTTCCGCGCCACACCTTGCANAACGCTGCCCAAGGCTGGAC 1650
551 TrpAspLysHisProSerHis 557
1651 TGGGACAGATCCCTCGCAC 1671

seq_name: gb_ba:PCYCBA
seq_documentation_block:
LOCUS PCYCBA 2629 bp DNA BCT 23-JAN-1992
DEFINITION P.denitrificans cyc A and cta dII genes for cytochrome c 550 and
cytochrome c oxidase beta subunit I.
ACCESSION Y07533.1 GI:45477
VERSION Y07533
KEYWORDS ctd dII gene; cycA gene; cytochrome c oxidase; cytochrome c-550.
SOURCE Paracoccus denitrificans.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Paracoccus.
REFERENCE 1 (bases 1 to 2629)
AUTHORS Raitio,M.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1990) M. Raitio, DEPT OF MEDICAL CHEMISTRY,
UNIVERSITY OF HELSINKI, SILTAVUORENRENNER 10, 00170 HELSINKI 17,
FINLAND
2 (bases 1 to 2629)
AUTHORS Raitio,M., Pilsa,J.M., Metso,T. and Saraste,M.
TITLE Are there isoenzymes of cytochrome c oxidase in Paracoccus
denitrificans?
JOURNAL FEBS Lett. 261 (2), 431-435 (1990)
MEDLINE 90184495
FEATURES Location/Qualifiers
source 1..2629

```



```

333 eutrhGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPro 349
|||||.....|
1871 TGACCACACAGGCGCTATTTCATGCTGGGAGACCATGACCATTGGCGTCC 1920
350 ThrGlyIleValIlePheSerTyrIleAlaThrMetTyrGlySerVal 366
|||||.....|
1921 ACCGGCATCAAGGCTTCTCTCGTGGATCGGACCATGTGGGGCGGCACAT 1970
366 IGIuPheIleSerProMetLeuTrrPalaPheGlyPheMetPheLeuPheT 383
|||||.....|
1971 CGAGTTCAAGACGCCGATCTCTGGGCTTGGCTTGGCTTCTGCTTCTGTTCA 2020
383 hrValGlyIleValThrGlyIleValLeuAlaGlnAlaGlyLeuAspArg 399
|||||.....|
2021 CCGTCGGGGCGGTGACCGCGCTGGTGGTGAACCGAGCGCGCTGGACCGG 2070
400 AlaIyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMetSe 416
|||||.....|
2071 GTCATACACGACACCTATTACGTGTGGCCGACATCCACATACGATGTC 2120
416 rLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetProL 433
|||||.....|
2121 GCTGGGGCGCGGTGTCGCGCATCTTCCGCGGGGTCTATTACTGATCGGCA 2170
433 ySPheSerGlyArgAlaPheProGluTrrPalaAlaLysLeuHisPheTyr 449
|||||.....|
2171 AGATGTGGGGCGCGCAATACCGGAAATGGCGCGCGCATGATTTCTGG 2220
450 ThrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuG 466
|||||.....|
2221 ATGATGTTTCATGCGCTGAAACCTGATCTTCTCCGCGACGATTCCTGGG 2270
466 yArgGlnGlyMetProArgTrrGlyIleAspTyrProGluAlaPheAla 483
|||||.....|
2271 CCGCGAGGCGATGCGCGCGCTATATCGACTATCCGCGTGGATGCGCT 2320
483 eutrrPalaIleValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPhe 499
|||||.....|
2321 ATTGGAAACACATCTCGTCATCGCGCCCTATATCTCTCCGCGCTCTTC 2370
500 LeuPhePheIleValIlePheValIleTyrThrLeuValAlaGlyArgG 516
|||||.....|
2371 CTGTTCTTCATCGCATCGTGTCTTACACGCTCTTCCGCGGACGCGCT 2420
516 uThrArgProAsnProTrrGlyLeuPheAlaAspTrrLeuGluTrrThrL 533
|||||.....|
2421 GAACGTCGCCAAGTACTGGAAGAGCATGCCGACGCTGGATGGAGCC 2470
533 eutProSerProProProlaHisThrPheGluThrLeuProlLysArgSer 549
|||||.....|
2471 TGGCCTGCGCGCGCGCGACATACCTTCGAGACCCCTGCCAAGCGCGAG 2520
550 AspTrrAspLysHisProSerHis 557
|||||.....|
2521 GACTGGGATCGCGACACGCGCAT 2544

seq_name: gb_ba:RSCOXSI
seq_documentation_block:
LOCUS RSCOXSI 2166 bp DNA BCT 09-APR-1992
DEFINITION R.sphaeroides gene for cytochrome oxidase subunit 1.
ACCESSION X62645
VERSION X62645.1 GI:46417
KEYWORDS cytochrome oxidase subunit 1.
SOURCE Rhodobacter sphaeroides.
ORGANISM Rhodobacter sphaeroides.
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Rhodobacter.
REFERENCE 1 (bases 1 to 2166)
AUTHORS Shapleigh,J.P.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1991) J.P. Shapleigh, University of Illinois,
Dept of Biochemistry, 505 S Mathews, Urbana IL 61801, USA

```

```

REFERENCE 2 (bases 1 to 2166)
AUTHORS Shapleigh,J.P. and Gennis,R.B.
TITLE Cloning, sequencing and deletion from the chromosome of the gene
encoding subunit I of the aa3-type cytochrome c oxidase of
Rhodobacter sphaeroides
JOURNAL Mol. Microbiol. 6 (5), 635-642 (1992)
MEDLINE 92204019
FEATURES
source
location/Qualifiers
1..2166
/organism="Rhodobacter sphaeroides"
/strain="Ga"
/db_xref="taxon:1063"
207..1904
/codon_start=1
/transl_table=1
/product="cytochrome oxidase subunit 1"
/protein_id="C0A44514.1"
/db_xref="GI:46418"
/db_xref="SWISS-PROT:P33517"
/translation="MADAAIHGEHEDHDFEFTRFMSSTNKHQIGVLYLFTGGLVGLIS
VAPRYVMRMELMARGVQFPCAEHLESLGVGFQSLSPSAVENCTNGHILANVITGH
GILMEFVILPALEGFGNFMPLHIGAPDMARPMNNLSYWLVA GSTLAVASLAP
GNGOISGIGWLYPLPLSTSESGYSTDLAIFAVHLSGASLIGAINMTTFLMRAP
GMTMKVPLFAMSIFVTAWLILALPYLAGAITMLLDIRNFSTFFQPSGGGPVLYQ
HILMFEGHPEVYIIVLPAGIVSHVITFAKRPFGYLPVYAMVALIGVLYVMAH
MYTAGLSLTOOSYFMATVIAVPTGKIRISWATMNGSIELKTPMLALGFLPLT
VGGGTGTVLSOASVDRYHDYVYVVAHFVMSLGAVFGIFASTGIGKMSRGQPE
WAGLHFMFMFVGCANLTFEPQHLGRGMRRTIDYPEARATNVESSIGALSTFASF
LFLGVTFYLSGARGVANNYMNEDATLEWLTLSPPHEFTQLPRRDWERAPAH"
BASE COUNT 326 a 746 c 639 g 455 t
ORIGIN
alignment_scores:
Quality: 2453.00 Length: 567
Ratio: 4.736 Gaps: 3
Percent Similarity: 91.358 Percent Identity: 78.660
alignment_block:
US-09-712-768-2 x RSCOXSI
Align seg 1/1 to: RSCOXSI from: 1 to: 2166
1 MetAlaAspPalaAlaIleHisGlyHisAspHisGlnGlyGlnGlyPh 17
|||||.....|
207 AAGCGCGAGCAGCAGCATCATGCGCAGAG...CACGACCGAGGGGCTT 253
17 ePheThrArgTrrPheMetSerThrAsnHisLysAspIleGlyLeuLeuT 34
|||||.....|
254 CTTCACCGCGCTGTTTCATGTCGACGACACACAGACATCGGCGTTCTCT 303
34 yrrLeuValAlaIleGlyValIleGlyPheIleSerValLeuPheThrVal 50
|||||.....|
304 ATCTCTTTCACCGGGGCGCTCGTGGGTGATCTCGGTCTTACCGCTTC 353
51 TyrMetArgLeuGluLeuMetAspProGlyValGlnTyrMetCysLeuG 67
|||||.....|
354 TACATCGGATGAGCTCATGCGCGCGCGGTGCGATCATGTGCGCGCA 403
67 uGlyAlaArg.....LeuIleAla 74
|||||.....|
404 ACATCTGGAAATCGGCGCTCGGAAAGGCTTCTCCAGTCGCTGGCGCT 453
74 sPalaSerGlnThrCysThrAlaAsnGlyHisLeuTrrAspValMetVal 90
|||||.....|
454 CGGCGGTGGAACACTGACCCCGAGCGCATCTGTGGAACGTCATGATGC 503
91 ThrTyrHisGlyIleLeuMetMetPhePheValGlyIleProAlaLeuPh 107
|||||.....|
504 ACCGGCAGAGGATCTCATATGTTCTTGGTCGATTCGCCGCTTT 553
107 eGlyIlePheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspM 124
|||||.....|

```

554 CGCGGCTTCGGCAACTATTTTCATCGCGCTGCACATCGGCGCGGACA 603
 124 evlAahepProArmetAnasInleuSerPheTrpLeuPheIleAaIagly 140
 TGGCTTCCTCCGCGATGAACACCTCTCTACTGGCTCATGTCCGGGG 653
 604 TGGCTTCCTCCGCGATGAACACCTCTCTACTGGCTCATGTCCGGGG 653
 141 ThrAlaMetGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLe 157
 654 ACCTGGCTTCGCGCTCGCTCTCTCGCGCGGCGGACGCGGACCT 703
 157 uGlySerGlyValGlyTrpValLeuTrpProLeuSerThrArgGlyA 174
 704 GGGCTTCGCGATCGCTGGCTGCTCATCGCGCTCTCTACCTCGCAAT 753
 174 IaglyTrSerMetAspLeuValAlaIlePheAlaValAlaHisLeuSerGlyAla 190
 754 CGGCTTCCTTCACCGACCTCGCATCTCGCGGCTGCACCTGCGGGCGCC 803
 191 SerSerIleMetGlyAlaIleAsnMetIleThrTrpPheLeuAsnMetAr 207
 804 TCCTCGATCCCTCGCGCGATCAACATGATACGACCTCTCTGAACATGGG 853
 207 gAlaProGlyMetThrLeuHisIleValProLeuPheSerTrpSerIleP 224
 854 CGCACCGCGCATGACATGCACAAAGTGGCGCTTCGCGCTGTCGATCT 903
 224 heIleThrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAlaGlyAla 240
 904 TCGTACACGCGCTGGCTGATCTGCTGCGCGCTGCGCGCGCGCGCC 953
 241 IleThrMetLeuLeuThrAspArgAsnPheGlyThrTrpPheAsnPr 257
 954 ATTACCATGCTGCTGACCGACCGGACCTTGGCACACCTTTCTCCAGCC 1003
 257 oAlaGlyGlyAspProIleLeuTrpGlnHisIleLeuTrpPhePheG 274
 1004 CTGCGGGGCGGCGCGCGCTGCTCTACACACATCTCTGCTTCTTCG 1053
 274 IyHisProGlyValTrpIleIleIleLeuProGlyPheGlyIleIleSer 290
 1054 GCCACCGCGAGGTCTACATCATCGCGCGCGCTTGGCATGTCAGC 1103
 291 HisValSerTrpPheSerIleValProValPheGlyTrpLeuProme 307
 1104 CACGTCATGCGACCTTCCCAAAAGCCGATCTTGGCTTTGGCGCAT 1153
 307 tValTrpAlaMetValAlaIleGlyValLeuGlyPheValValTrpAla 324
 1154 GGTCTATGCGATGGTGGCGATGCGGCTGCTGCTGCTGCTGCGGCC 1203
 324 IeHisMetTrpThrValGlyMetSerLeuThrGlnSerTrpPheMet 340
 1204 ACCACATGATACACCGCGCTGAGCCTCACCCACAGAGCTACTTATG 1253
 341 LeuAlaThrMetValIleAlaValProThrGlyIleIlePheSerTr 357
 1254 ATGGCAGCATGATGATTCGACATGCCACCGCATCAAGATCTTCTCTG 1303
 357 pIleAlaThrMetTrpGlyGlySerValGlnPheLeuSerPheMetLeu 374
 1304 GATCGGACCATGTGGGCGGCTCGATCGATCAAGAGCGGATGCTCT 1353
 374 rPlAlaPheGlyPheMetPheLeuPheThrValGlyGlyValTrpGlyIle 390
 1354 GGGCGCTCGGCTTCCTCTCTCTTACCGTGGGCGGCTCACCGCATC 1403
 391 ValLeuAlaGlnIleAspArgAlaTrpHisAspThrTrpVal 407
 1404 GTCGTGAGCAGGCGAGCGTGCACCGCTATTATCAAGACACTACTATGT 1453
 407 ValAlaHisPheHisTrpValMetSerLeuGlyAlaIlePheAlaIleP 424
 1454 CGTGGCGCACTTCATTTATGTATGAGCTCGGCGGCTTCTCGCATCT 1503

424 heAlaGlyIleTrpThrPheTrpMetProIyPheSerGlyArgAlaPhePro 440
 1504 TCGAGATCTTACTTCTGGGATCGGCACATCTCGCGCGCATATTCG 1553
 441 GluTrpAlaAlaIleValLeuHisPheTrpThrPhePheIleGlyAlaAsnVa 457
 1554 GAAATGGCGCGGAAGCTCATTTCTGATGATGATTTCTGGGCGCAACT 1603
 457 ThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProArgArgT 474
 1604 CACCTTCTTCCGACACTCTCTCGCGCGCATGCTCTGCTGCTGCT 1653
 474 YrIleAspTrpProGlyValAlaPheAlaLeuTrpAsnIleValIlePheVa 490
 1654 ACATGCACTTCCGAGGCGCTTCCGACCTGGAACCTTCTGCTGCTG 1703
 491 GlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleValIlePheVa 507
 1704 GCGCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1753
 507 ITrpThrLeuValAlaGlyArgGluTrpArgProAsnProTrpGlyG 524
 1754 CTACAGCTG...AGCGGCGCGGCGGAGGAGCGGACAACTACTGAGACG 1800
 524 IupheAlaAspThrLeuGluTrpThrLeuProSerProProAlaHis 540
 1801 AACATCCGACACCTGAGTGCAGCTGACCTCTCTCTCTCTCTCTCT 1850
 541 ThrPheGluTrpLeuProIyArgSerAspTrpAspIyHisProserH 557
 1851 ACGTTCGAGCACTTCCAAAGGGAAGACTGGAGACGCGCGCCGCCA 1900
 557 s 557
 1901 C 1901
 seq_name: gb_da:PDCCX2
 seq_documentation_block: 3211 bp DNA BCT 07-OCT-1996
 LOCUS PDCCX2
 DEFINITION P. denitrificans COI gene.
 ACCESSION X05829
 VERSION X05829.1 GI:994806
 KEYWORDS COI gene; cytochrome c oxidase; cytochrome oxidase a3;
 unidentified reading frame.
 SOURCE
 ORGANISM
 Paracoccus denitrificans.
 Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 Paracoccus.
 1 (bases 1 to 3211)
 Raitio,M., Jalli,T. and Saraste,M.
 Isolation and of the genes for cytochrome c oxidase in Paracoccus
 denitrificans
 EMBO J. 6, 2825-2833 (1987)
 2 (bases 1 to 3211)
 Saraste,M.
 Direct Submission
 Submitted (04-JAN-1991) Saraste M., EMBL, Meyerhofstrasse 1,
 Heidelberg, Germany
 revised by [3] MAT
 3 (bases 1 to 3211)
 Saraste,M.
 Direct Submission
 Submitted (20-SEP-1995) Saraste M., EMBL, Meyerhofstrasse 1,
 Heidelberg, Germany
 On Sep 21, 1995 this sequence version replaced gi:45474.
 COMMENT
 FEATURES
 source
 1..3211
 /organism="Paracoccus denitrificans"
 /db_xref="taxon:266"
 /map="cox2 locus"
 218..222
 RBS


```

363 GLYGIServValGluPhelySerPrometLeuTPAlaPheGlyPhe 379
1857 GGGGGCTCGATTCAGATTCAGACGCCGATGCTCTGGCCCTGGCC... 1901
379 tPheLeuPheThrValGlyValThrGlyValLeuAlaGlnAlaG 396
1902 .TTCCGTTCACCGTCGGCGCGCTGACCGGGGTCGATCGACCGCAGGGGT 1950
396 tLeuAspArgAlaAlaThrAspThrTyrTyrValAlaAlaHisPheHis 412
1951 CCTGGACCCCGTTTATACAGACACCTACTATATGCGGGCGCATTTTCAC 2000
413 TyrValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPh 429
2001 TATGTATGAGCCCTTGGCCCTCTTCGCGATCTTTCAGAGCACCTATT 2050
429 eTyrMetProLySPheserGlyArgAlaPheProGluTyrPAlaAlaLysL 446
2051 CTGGATCGGAGATGTCGGCGGCAATACCGGAAATGGCGGCGGCGCAGC 2100
446 eHisPheThrPhePhePheIleGlyAlaAsnValThrPhePheProGln 462
2101 TGCATTTCTGATGATGATTCATCGCTCGAACCCTGATCTTCTCCGACG 2150
463 HisPheLeuGlyArgGlnGlyMetProArgTyrIleAspTyrProG 479
2151 CACTTCTCGGGCGCGGCGGATGCGCGCGCGCATATTCACACACCGGT 2200
479 uAlaPheAlaLeuTyrAsnLysValSerSerTyrGlyAlaPheLeuAlaP 496
2201 CGAATCTCGACTGGAACAACATCTGTCGATCGCGCGCTATATCTCT 2250
496 heAlaSerPheLeuPhePheIleValIlePheValTyrThrLeuValAla 512
2251 TCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2300
513 GtYArgArgGluThrArgProAsnProTyrGlyGluPheAlaAspThrLe 529
2301 GGCAGAGCCCGGAGACGTCGCACTACTGGAACAGACATCGGACAGCT 2350
529 uGluTyrThrLeuProSerProTyrProAlaHisThrPheGluThrLeuP 546
2351 GGAATGAGCCCTGCTCGCGCGCGCGCGGACATACCTTCGAGACCTGCG 2400
546 tOLYSArgSerAspTyrAspLysHisProSerHis 557
2401 CCAAGCCCTGAGGACTGGATGCGGACACAGGCGCAT 2435

seq_name: gb_ba:SME591785

seq_documentation_block:
LOCUS SME591785 286550 bp DNA BCT 16-AUG-2001
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 4/12.
ACCESSION AL591785 AL591688
VERSION AL591785.1 GI:15073719
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
Rhizobiaceae: Sinorhizobium.
1 (bases 1 to 286550)
REFERENCE
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boisard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
From the Cover: Analysis of the chromosome sequence of the legume
symbiont Sinorhizobium meliloti strain 1021
Proceedings of the National Academy of Sciences of the United
States of America. 98 (17), 9877-9882 (2001)
PUBMED 11481430
JOURNAL 2 (bases 1 to 286550)
REFERENCE Gouzy,J.
AUTHORS

```

```

TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
COMMENT RU Consortium
MELILO RU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP227, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.gouzy@oulouise.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.

FEATURES
source
1..286550
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
70..813
/gene="gst11 OR SMC00916"
/EC-number="2.5.1.18"
/function="miscellaneous"
/note="Product confidence : putative
Gene name confidence : putative
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not experimental
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/transl_table="MADLSFPTTARWPAKPNPDIIQVLSLPTNGKVSJALEELGIA
YEARIISFDTEWKSPEFLSNPGRIPALIDPGPGKPIGLESGALILYLAETG
KILPADARREYETLCWVFMGCGIGFPGFKAADKVNANPYVQRYDEAKR
LLGVLEAREBROWLMGSRNPTSPRSRPTGKMGKRGSSVAPRCSTIRASRMWMTGI
AAWALRRRRMRKRSRARSRSRSLPRAGD"
588..779
/gene="SMC00915"
/EC-number="779"
/function="miscellaneous: hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence="not experimental
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/transl_table="MADGFEYTIADIATYPMVGARKFYGAELVDYKSPFNMDWV
DRGLARPAOKGMEIPRE"
867..2084
/gene="SMC00914"
/EC-number="2084"
/gene="SMC00914"
/function="small molecule metabolism"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11

```

```

gene
CDS
        /gene="SMc00910"
        /locus_tag="6004..6852"
        /function="miscellaneous; hypothetical/global homology"
        /note="Product confidence: hypothetical
        gene name confidence: hypothetical
        predicted by Codon usage
        predicted by Homology
        predicted by Framed"
        /codon_start=1
        /transl_table=1
        /evidence=not_experimental
        /product="CONSERVED HYPOTHETICAL PROTEIN"
        /protein_id="CAC45367.1"
        /db_xref="GI:15073726"
        /translation="MAVRINSFRIEAGRDVYLCDWGYLHNGIOAFASCEALAEAR
        AOGIVLITISPRHFGVTVQIRGLGVDEAYDRIVTSGDVTOLIAAADRLEFFIG
        ADRLDPLEIGLSTETIVSADAEETVCAGEYDETEPERHATLTGLAKRRIPTICAN
        PDLVVERGHRILPCAGAIARLYEELGGEARIAGKPYIAYRAALAEAKAAGARDLSR
        VALIDGMPITDVKGAODAGFDLLYSAGIHAGCYMHESPTDEARLAFRLQNGAPKKK
        WMPRIA"
gene
        /gene="ribp OR SMC00909"
        /locus_tag="6866..7849"
        /gene="ribp OR SMC00909"
        /EC_number="2.7.1.26"
        /EC_number="2.7.7.2"
alignment_scores:
        quality: 1898.50
        ratio: 4.039
        Percent Similarity: 85.299
        Percent Identity: 63.158
Alignment_block:
US-09-712-768-2 x SMC591785 ..
Align seg 1/1 to: SMC591785 from: 1 to: 286550
3 ASPAlAlAlleHISGLyHIS...ASpHIS...HISGLySGInGLyph 17
||||| ::: ||||||| ||||| ||::: ||
126428 GATCATATCCGATCATGCGCAGCGGACACGCGGCGACGCGTGCATCTT 126477
17 ePhetHrArgTrpHemeSerThrAsnHisIysaspriLeIyLeuMet 34
||||| ||||||| ::: ||||||| ||||| ||||| ||||| |||||
126478 CTTCACAGCGCTGGTTCCTTCAACACACACAGGACATCGGTAAGCTCT 126527
34 yrlEuuValAlaAlaGlyValAlaGlyPheIleSerValIleuPheThrVal 50
||||| ||| :::: ||| :::: ||| :::: ||| :::: |||
126528 ACCTGATCTTCGCGATCATCGCGGCGCATCAACGCGGCGACGCTGGCTC 126577
51 TyrMetArgLeuGluMetLaspProGlyAlaGlnTyrMetCysLeuG1 67
||||| ||||| ||||| :::: ||||| ||||| ||||| |||||
126578 TTTCATCGCGCGCCAACTGCAGGAGCGGCGCATCAAG.....ATCTTCCA 126621
67 uGly...AlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnG 83
||||| ||||| :::: ||||| :::: ||||| :::: |||||
126622 CGGTCGCGCCACAGATGCTATAGCTTTTCGAGGGTATGCTGCCATTCGAGC 126671
83 ly.....HisLeuTrpAsnValMetValIhTrpHisGlyIleLeuMet 97
||||| ||||| :::: ||||| ||||| :::: |||||
126672 GCGCGAAGCAACATGTTCAACGCTTCCACGACCGCTCACGCGGTGATCATG 126721
98 MePhePheValAlaGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrlE 114
||||| :::: ||||| :::: ||||| :::: ||||| :::: |||||
126722 ATCTTCTTATGTCATGCCGCGCTGAATGGCGGCTTCGCAACTGATGAT 126771
114 uMetProLeuGlnIleGlyAlaIleProAspMetAlaPheProArgMetAsnA 131
||||| :::: ||||| :::: ||||| :::: ||||| :::: |||||

```

126772 GGTCCGATCATGATGCGCGGACATGGCGCTCCCGCATGACAA 126821
 131 sLeuSerPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSer 147
 126822 ACATTTCCTTGGGTGATGATGCGCGCTCTCTGCTGCTGCTTGG 126871
 148 LeuPheAlaProGlyLysPglYsInLeuGlySerGlyValGlyTrpAla 164
 126872 ATGTGTGGAGAGTCTCTGAGGCGCTATGGCGGCGGCGGCGGAGAC 126921
 164 IleuTrpProLeuSerThrArg...GluAlaGlyTrpSerMetAspL 180
 126922 GATCTATTCGCGCATCTCTGACATCCGCGCATGCCCGGCGCGGATGATC 126971
 180 euAlaIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAla 196
 126972 TCCGATCTCTGCGGCTGACATGCGGCGCGCTCTGATCTCCGCGGCG 127021
 197 IleAsnMetIleThrPheLeuAsnMetArgAlaProGlyMetThrIle 213
 127022 ATCACTTCTACCAACCATCTCAACATCCGCGCTCCGCGGATGACCT 127071
 213 uHisLysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeu 230
 127072 GCACAGATGCGCGCTCTTGGCTGCTGCTGCTGATCACCGCTCTCTGC 127121
 230 IleLeuLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuTr 246
 127122 TCCGCTCTGCGGCGGCTCTGCGAGCGGATGACCAACATGCTGCTGACG 127171
 247 AspArgAsnPheGlyThrThrPheAsnProAlaGlyLysLysAsp 263
 127172 GACCGCACTTCCGCGACGACCTTCTGCGCGGAGGCGGCGGCGACCC 127221
 263 oIleLeuTrpGlnHisIleLeuTrpPheGlyHisProGlyValTrp 280
 127222 GATCTCTGCGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127271
 280 IleIleIleLeuProGlyPheGlyIleIleSerHisValAlaSerTrpPhe 296
 127272 TTCGATCTCTGCGGCGGCTCTGCGACGACCAACATGCTGACCTTTC 127321
 297 SerLysLysProValPheGlyTrpLeuProMetValTrpAlaMetValAl 313
 127322 TCCGCGCAAGCGCATCTGCTGATCTGCGGATGCGCCATGCGTGGC 127371
 313 aIleGlyValLeuGlyPheValValTrpAlaHisHisMetTrpTrpValG 330
 127372 GATGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127421
 330 LysMetSerLeuThrGlnGlnSerTrpPheMetLeuAlaTrpMetValIle 346
 127422 GCATGTGCTCTGACACAGCGCTACTTCTGCTTCTGCGGACGATGGTATC 127471
 347 AlaValProThrGlyLysIlePheSerTrpIleAlaThrMetTrpG 363
 127472 GCGGCGCGGCGGCTGATGATCTCTGCTGCTGCTGCTGCTGCTGCTG 127521
 363 yGlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetP 380
 127522 CGGCTGCGATCGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 127571
 380 HeLeuPheThrValGlyValThrGlyIleValLeuAlaGlyAlaGly 396
 127572 TCTGTTCACGCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 127621
 397 LeuAspArgAlaTrpHisAspThrTrpValValAlaHisPheHisTrp 413
 127622 CTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 127671
 413 rValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyLysTrpPhe 430
 127672 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 127721

430 yrmMetProLysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeu 446
 127722 GATTCGCGAAGATGACGCGCTACATGATCTGAGATTTCTTCCGCAACTG 127771
 447 HisPheTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnH 463
 127772 CATTTCTGGGATGATGCTGCGCGGTGAACTGATCTCTTCCCGGACGA 127821
 463 sPheLeuGlyArgGlnGlyMetProArgTrpTrpIleAspTrpProGlu 480
 127822 TTTCTCTGCGGCTGCGGAGCATGCGCGCGCTTACATGACATATCCGAGC 127871
 480 laphAlaLeuTrpAsnLysValSerSerTrpGlyAlaPheAlaPhe 496
 127872 CTTTCCGCGCGCTGAGACATGTTGCTTCTTACGCTCTTACATCGCGCA 127921
 497 AlaSerPheLeuPhePheIleValIlePheValTrpThrLeuValAlaG 513
 127922 GTGCGCGTGTGATCTTCTCT...TTCTGTGCTGCGCGAAGCTTTCCGGA 127968
 513 yArgArgGluThrArgProAsnProTrpGlyGluPheAlaAspThrLeuG 530
 127968 GAAGGCGCTGCGGCGGATACCCCTGCGGCGAAGCGCCACACAGCTGG 128018
 530 lutyTrpThrLeuProSerProProAlaHisThrPheGluThrLeuPro 546
 128018 AATGGCAGCTGTCTTCCGCGCGGCTTCCACCAATGGAGAGAGCTCCG 128068
 547 Lys 547
 128069 CCG 128071

seq_name: gb_ba:R1COXA
 128069 CCG 128071
 seq_documentation_block:
 LOCUS R1COXA 1707 bp DNA BCT 25-APR-1994
 DEFINITION R. leguminosarum coxa gene.
 ACCESSION X74341
 VERSION X74341.1 GI:396698
 KEYWORDS cytochrome c oxidase; cytochrome c oxidase subunit I.
 SOURCE Rhizobium leguminosarum.
 ORGANISM Rhizobium leguminosarum
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Rhizobium.
 REFERENCE 1 (bases 1 to 1707)
 AUTHORS Gabel, C.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUL-1993) C. Gabel, The Johns Hopkins University,
 Dept. of Biology, Charles and 34th Sts., Baltimore MD 21218, USA
 REFERENCE 2 (bases 1 to 1707)
 AUTHORS Gabel, C., Bittinger, M.A. and Maier, R.J.
 TITLE Cytochrome a3 gene regulation in members of the family
 Rhizobiaceae: comparison of copper and oxygen effects in
 Bradyrhizobium japonicum and Rhizobium tropici
 JOURNAL Appl. Environ. Microbiol. 60 (1), 141-148 (1994)
 MEDLINE 9416196
 FEATURES
 source
 1. 1707
 /organism="Rhizobium leguminosarum"
 /strain="USDA 27"
 /db_xref="taxon:384"
 91. 1707
 /gene="coxa"
 91. 1707
 /gene="coxa"
 /EC_number="1.9.3.1"
 /codon_start=1
 /transl_table=11
 /product="cytochrome-c oxidase, subunit I"
 /protein_id="CAA52388.1"
 /db_xref="GI:396698"
 /db_xref="SWISS-PROT:Q08855"

gene
 CDS

/translation="MATAAAGEHAEDHDEHAPPTGRSTNHKIDIGTLLYIFAI
 IAGVIGAAASLAIRAEIAMPVGEYFHTLVYFVTHVIFEMVPMAMGGGNW
 FLPLMIGADPMAPRMNNTSEMLPASPGLLMTYVEEPEPANAGAGMTMYVPLSS
 SGHPAPVDAIPESLHAGASITLGINFTITLNNRAGMLHMKDPLFAMSVLTAR
 LLLSLPVLGATITMLTDNRNCTIGTFEAREGGEDLLYOHLEWFGHEVYLLILPGF
 GMSHITSTPSKRPVFCYIGMYAMALIGLGFVVAHMYTVGMDLEAVFSAVM
 IIAVPTAFIKESMIATMWSGLEAFAPMLMALAFLELFVGGVGVLANSLDRLVH
 DTYVVAHFHYVLSLGAIFPAFAGWYFPMKSGMYNETLEAHFMIIFISVNLIF
 PEHFLIGSMPRRYIDYDPDAFAGMNLVSSIGSYISGFVLLFTCYVDAFAKNVPGD
 NPMGAGATLLEWTLPSPPVHEFEVLPVE"

BASE COUNT 277 a 537 c 479 g 414 t
 ORIGIN

alignment_scores:

Quality: 1873.50 Length: 542
 Ratio: 4.173 Gaps: 4
 Percent Similarity: 82.841 Percent Identity: 64.207

alignment_block:

US-09-712-768-2 x RLCOA ..

Align seg 1/1 to: RLCOA from: 1 to: 1707

7 HisGlyHisAspHisGlyHisGlyGlyPhePheThrArgTrpPhe 23
 |||||
 133 CATGGCCAGATGAAACACACGCCGCCG.....ACGGCTGGGCCCG 173
 23 tSerThrAsnHisLysAspIleGlyLeuLeuTyrrLeuValAlaAlaGly 40
 |||||
 174 CTCGACCAACCATAGATATCGGTACCGTACCTGATCTTGGCTATCA 223
 40 aValGlyPheIleSerValLeuPheThrValTyrrMetArgLeuGlyLeu 56
 |||||
 224 TCCGAGGAGATCGGGGACCATGCTGGCCATCCCGCGGAACCTG 273
 57 MetAspProGlyValGlnTyrrMetCysLeuGlyAlaArgLeuIleAl 73
 |||||
 274 ATGACCCGGGGCGGTGAGTCTTC..... 297
 73 aAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMet 90
 |||||
 298CATACACCATCTGTACACGATATGCG 325
 90 aLhrTrpHisGlyIleLeuMetMetPhePheValGlyIleProAlaLeu 106
 |||||
 326 TTACCTCGCACGGCGTTATCATGATCTTTCATGGTCATGCGGCTATG 375
 107 PheGlyGlyPheGlyAsnTyrrLeuMetProLeuGlnIleGlyAlaProAs 123
 |||||
 376 ATGGGGGGCTTCGGCAACTGTTCCGCCCTGATGATGCGGCCCGCA 425
 123 pMetAlaPheProArgMetAsnAsnLeuSerPheTrpLeuPheIleAlaG 140
 |||||
 426 TATGGCTTCCCGCGATGAAATATCTGCTTGGCTCCGCCGCTT 475
 140 LyrThrIleMetGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGln 156
 |||||
 476 CGTTCGGCTGCTGCTCATGTCACCTTCGACAGGAGGGAACCGGGTGC 525
 157 LeuGlySerGlyValGlyTrpValLeuTyrrProProLeuSerThrArg.. 172
 |||||
 526 AATGGCGGGGCTGGCGGAGATGATGTCGCCGCTCATCATCAGCAG 575
 173 GlnAlaGlyTyrrSerMetAspLeuAlaIlePheAlaValHisLeuSerG 189
 |||||
 576 CCATCCGGCGCCCGCGGTGATCTGCAATCTTTCGTCGCAATATGCGG 625
 189 LysAspSerIleMetGlyAlaIleAsnMetIleThrThrPheLeuAsn 205
 |||||
 626 GCGCTTCGTCGATCTCGGTGCGATACCTTCATCCACGATCCCTGAAC 675
 206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSe 222

|||||
 676 ATGGCGGCTCCCGCATGACGCTCCCAAGATGCCCCCTGTGGCCGGTC 725
 222 rIlePheIleThrAlaTrpLeuIleLeuLeuAlaLeuProValIleAlaG 239
 |||||
 726 GGTTCGATACCGCATTCCTGCTGCTTCGTCATCCCGGATCTCCGCCG 775
 239 LysAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePhe 255
 |||||
 776 GCCCATTCACCATGCTGTGACCGACGCAATTCGGGACGACCTTCCTT 825
 256 AsnProAlaGlyGlyLysAspProIleLeuTyrrGlnHisIleLeuTrpH 272
 |||||
 826 GCGCCGGAAGCGCGGTGATCCGCTGCTACACACGACCTCTTCGGTT 875
 272 ePheGlyHisProGlyValTyrrIleIleIleLeuProGlyPheGlyIleI 289
 |||||
 876 CTTCGGCATCCGGAAGTCTCATCTTCGATCCGCCGCTTCGCGCATGA 925
 289 IeSerHisValIleSerThrPheSerLysProValPheGlyTyrrLeu 305
 |||||
 926 TCAGCCCATCATCTCGACCTTCGCGCAAGCCGGTATTCGGTACATC 975
 306 ProMetValTyrrAlaMetValAlaIleGlyValLeuGlyPheValValTr 322
 |||||
 976 GGCATGATATACGATATGCTGCTGATCCGCGGCTGGGGTGTGTGCTCG 1025
 322 pLHisHisMetTyrrThrValGlyMetSerLeuThrGlnGlnSerTyrr 339
 |||||
 1026 GCGCCACCATGATATCTGCGCATGATGATGATACCGGAAGGCTATT 1075
 339 hMetLeuAlaThrMetValIleAlaValProThrGlyIleLysIlePhe 355
 |||||
 1076 TCGTTCGGCGCACCATGATCATTCGTGTGCCGACGGATCAAGATCTTC 1125
 356 SerTrpIleAlaThrMetTrpGlyLysSerValGluPheLysSerProme 372
 |||||
 1126 TCATGATTCGACCATGTCGGGTGTAGCATGAAATTCGACACCCCAT 1175
 372 tLeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyGlyValThr 389
 |||||
 1176 GCTGTGGGCGCTGGCGTTATCTTCTGTTACGCTTGGCGCGGTATACCG 1225
 389 LyrIleValLeuAlaGlnAlaGlyLeuAspArgAlaTyrrHisAspThrTy 405
 |||||
 1226 GGGTGTCTCTCGCTAACCGACGTCGATCGGCTCCACGATACCTTAC 1275
 406 TyrValValAlaHisPheHisTyrrValMetSerLeuGlyAlaIlePheAl 422
 |||||
 1276 TACGTCGTTGCTCATTTCCACATGCTCTGTGCTGGGGCTATTTTCGC 1325
 422 aIlePheAlaGlyIleTyrrPheTyrrMetProLysPheSerGlyArgAla 439
 |||||
 1326 CATCTTCGCGGCTGTACTACTGTTCGCCAAGATGTCGGGGATATATGT 1375
 439 hProGlyTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGlyAla 455
 |||||
 1376 ACAACGAACCCCTGGCGGACGACACTTCGTCGATGATTCATGCGGCTC 1425
 456 AsnValIleThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetPro 472
 |||||
 1426 AACCTGATCTTCTCCCGACGACTTCTCGGCAATTCGGGAAACCGCGC 1475
 472 gArgTyrrIleAspTyrrProGlnAlaPheAlaLeuTrpAsnLysValSer 489
 |||||
 1476 CCGGATATTCGATACCGCGAGCGGTTGGCGGCTGGAATTCGTTTCAT 1525
 489 eTyrrGlyAlaPheLeuAlaPheAlaSerPheLeuPheIleValIle 505
 |||||
 1526 CCATCGGCTGTACATCTCGGGTTCTCGGCTTCGCTTCATCTACTGCG 1575
 506 PheValTyrrThrLeuValAlaGlyArgArgGlyLysThrArgProAsnPro 522
 |||||


```

1576 ...GCTACGATCGCTTGGCAAGACGTCCGGAGCGGACACATCCCTG 1622
522 pgiygluPhueAlaaspThrlengluThrlProserProprioProA 539
||||| ||| |||||||||||||||||||||||||||
1623 GGGCGGGGGCGGACACCCCTCGAATGAGCGCTCCCTCGCGCGCGG 1672
539 laH1sthrPhegluThrlLeuProLys 547
::||| |||||||||||||||
1673 TCACGAGTTCGAACTCTGCCCCG 1698

seq_name: gb_ba:AE008010

seq_documentation_block:
LOCUS AE008010 10835 bp DNA BCT 14-AUG-2007
DEFINITION Agrobacterium tumefaciens strain C58 circular chromosome, section
68 of 254 of the complete sequence.
ACCESSION AE008010 AE007869
VERSION AE008010.1 GI:15155740
KEYWORDS
SOURCE Agrobacterium tumefaciens.
Agrobacterium tumefaciens
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 10835)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
2 (bases 1 to 10835)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
location/Qualifiers
1..10835
/organism="Agrobacterium tumefaciens"
/strain="C58"
/db_xref="taxon:358"
complement(118..642)
/gene="AGR_C_1396"
complement(118..642)
/gene="AGR_C_1396"
/note="(AJ242592) hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1396p"
/protein_id="AAK86575.1"
/db_xref="GI:15155741"
/translation="MRLSPLARFLAAGLAIIPAAAAAQAQPPGTPKSTHGAMSV
TCDPAGASEBOCALMONYIADDPREGVLSVILTKADRSRLRLAPLGLVKDGM
ELIYDNNNGRAYFTFCFSEGCVEVDIDDELKTLRAGKNVFPALRESVDQDRGIP
IELSGFASGIDALP"
988..1938
/gene="AGR_C_1397"
988..1938
/note="(AJ242592) cytochrome c oxidase subunit II"
/codon_start=1
/transl_table=11
/product="AGR_C_1397p"
/protein_id="AAK86576.1"
/db_xref="GI:15155742"
/translation="MDSASYIDDOTREGGVTNRITAAAGMTCLTAVGAHADDP
VHMOMQOEAPRIMHEIRMEFQDTLPIVPTLFLVALLITVAVKFAKNPVAASKT
SHNTALEVWTLPVLTLLFLAPPSNLNAOITOPENDITLKATQIMWSYEKA
AEGAEPLSPDSYLLKQDRAAKEDKARIPRLAVDENKVPYGVKTVRLITAAPTD
VIAHAFMAPGVKIDAVPGRINLNTWPKPEKEGLLYGQCSLCKGDHAFMPLATRVSE
OQYNTMHAASDLNGLNANRALMASVDGAPRTVDVAANETN"
2549..3664
/gene="AGR_C_1399"
2549..3664
/gene="AGR_C_1399"

gene
CDS
/feature="AGR_C_1399"

gene
CDS
/feature="AGR_C_1399"

gene
CDS
/feature="AGR_C_1399"

gene
CDS
/feature="AGR_C_1399"

/note="CYTOCHROME C OXIDASE POLYPEPTIDE I (CYTOCHROME AA3
SUBUNIT 1)"
/codon_start=1
/transl_table=11
/product="AGR_C_1399p"
/protein_id="AAK86577.1"
/db_xref="GI:15155743"
/translation="MDLAFSLHVGASSILGAINETITLNMRAQMTLHKMPLFAM
SVLVTRFILLISLPVLAGITMLTDNRNGTAFSPSEGGDILYQHLFWFGHPREV
LILPFGITSHLVSTFSKRPYGVGLGMAVAMVAIGANGFIVMAHMYTGLSLEAOR
YFVFATMVIAVPTGIKIFSMVIAIMWGSLFSTPYMAIGFTFLTVGSGYVQCLANA
GIDRSLSHDTYVVVAHFHYVLSGAVAFPAQWYWPKITGYMNEFTCKLRFWMEFI
GVNLIFFPHQFLGIAQMPRRYIDPDYAPRAGMNVSSYGSYISAVAVGLFLFCVWBAFA
KRLIAGNPMGEGATTLFEMQLSPPYHQMQLPRIR"
4054..5007
/gene="AGR_C_1402"
4054..5007
/note="(X89566) heme O synthase"
/codon_start=1
/transl_table=11
/product="AGR_C_1402p"
/protein_id="AAK86578.1"
/db_xref="GI:15155744"
/translation="MTVIDRDMGAESSELSFAGRDYFELIKPVMNLVYPTAFAG
LVLAPEINPILGLIILICAVAGASGALNMVYDADIVAMTRAKRPIPSGRILAPR
EALAFGLTISAFSVILGLIIVNMFSGALAFTEFFAVVYTMILNKRSTQNTVIGGA
GAPPNMGCACMGVSDVILFLIPLMTPAHFPAALAFKRDYAGVIGPMPVPA
GERSTKQMTVYVLPRAAIVAPFFTGISAAGVIFAANVAFVYCSDFRMRMPCGD
EKMLPAKMFAYSVLYLFLFSGLLADHFAIKAVISGL"
5167..5784
/gene="AGR_C_1406"
5167..5784
/note="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
/codon_start=1
/transl_table=11
/product="AGR_C_1406p"
/protein_id="AAK86579.1"
/db_xref="GI:15155745"
/translation="MANMEAEATAGASQRYNSRISLVLCVFECAMIGMAVAAVPLYS
LFCVTVNGYQTVRQVSDVILDKTINVFEDANTSGLMNDPRVDKVKPKIGETI
KVTIKATINRSVATTTGTAIVNPTMEGACFENKVECCPFTLLOPFTLEMPVFEV
DPDIATAREKSVHTLSTFTFPAKTERVASLPVTEGDSKL"
5846..6721
/gene="AGR_C_1407"
5846..6721
/note="(AJ242592) cytochrome c oxidase subunit III"
/codon_start=1
/transl_table=11
/product="AGR_C_1407p"
/protein_id="AAK86580.1"
/db_xref="GI:15155746"
/translation="MADTHOKNHDYHIIDPSMPDLASIGAFIMTFGQVCMRYLSG
SEKLFGLARLANPWFYIGLVIVLVYVMAAMADTIIKENKPSGTEIVIDPMVILN
ASEVMEFYAMFATFDASLFPHEALQASRLLEPTGGOMKPGKIEIVIDPMVILN
LISGTVFAMHALLHNDKGLISGLATLVAGLYFSYQYVEYIINAPDEKNSIYA
TFFMATGFHGHFHFVGVTLVCLFRAIAGFTFKOHGFBAAMVIMHFVDVWLFLE
FAIYMGCGAPLHG"
6941..7369
/gene="AGR_C_1409"
6941..7369
/note="(AJ242592) hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1409p"
/protein_id="AAK86581.1"
/db_xref="GI:15155747"
/translation="MYSRFGQREREVLMSEPSNGNFAVPDPIVYGLKGGCPRGNGR
LFDGFLTPKPSACGLDYGADADGPAVFVMLLVGLVVGMLMWDORFAPYVMAH
VMLMLPEYIVSLVILRLKLGIMIALQYRNNSAEGRLDRE"

```

```

gene
7338..8114
/gene="AGR_C_1411"
CDS
7338..8114
/note="AGR_C_1411"
/note="(AF182952) Surfelt 1"
/codon_start=1
/transl_table=11
/product="AGR_C_1411p"
/protein_id="AAK86582.1"
/db_xref="GI:15155748"
/translation="MQAKGLIVNDVSHMKAORWVPAAPLVLAIALILGCTWQY
KRLVKEALMDIEERNASPATLSDEIAIVKSGEIEYKRVLSGTFDHTREHFA
THQGGTGYITPTPLADGRILFVNRGFVPEMKEAARPDGQVSGEVTITGLARAPL
VAKBSLLPDDNDIAKNIYFMKDLAMASSAEVPEPRVKKLFVDADNMPNGMFGQGV
TLIDLPNNHLGYAITWGLAALIVAGFAFRNGAKGE"
8173..9219
/gene="AGR_C_1414"
8173..9219
/gene="AGR_C_1414"
/note="(AE006203) LyCB"
/codon_start=1
/transl_table=11
/product="AGR_C_1414p"
/protein_id="AAK86583.1"
/db_xref="GI:15155749"
/translation="MNTAVSKRPPLTILRCGRFCAGVDRAIQIVLALAKAYPVVY
NHEIVHNRYVEGLEAKGALFVEELHEIPAEHRQPVVFAHGPKSVDPDAQRNLF
YLDATCPLEKSVHQAOMHQRHVVLIHAGHVEYIGTMGOLPEGTVAARNDL
GVEBPVRENLGVFTQTTLSYDADGVIARLOERPAIQAPADISICATYTNODAVK
QAAGCDLFIYVGAENSSSKRLVEVALRAGAKVAVVORASSTDMETIGDITVGS
AGASPEYIVDEIIEAFKARPDITLDLAVVEEIEHFLVNRKLSIELTIDDAFVNG
NASNALPRKATAG"
9312..10280
/gene="AGR_C_1416"
9312..10280
/gene="AGR_C_1416"
/note="HOMOSERINE KINASE (HK)"
/codon_start=1
/transl_table=11
/product="AGR_C_1416p"
/protein_id="AAK86584.1"
/db_xref="GI:15155750"
/translation="MAVYDTIDEDELNPLTOYDVGSLTSYKGIAGSVENSNLHTT
KDLILTLERKVKNDLPFLIGIMOHAKGSCPLPRKDGELIGELSGPALI
SLEGMWLRRPEAKHCEVCAALAAHMLASEGEITRPNALSVGMKVLMDKSEERAD
EVEKGLRETRPEIDYLAHMPKDLPGAVIHADLFQDNVEFLDELGLIDEFACND
FLAYDVSICTNACFEKDGAVNVTGKALLLEGYSVPVSEAELEALPLISRSALRF
FLRLYDMVLTTPAGALVKKDPLEYLRKLFHRTIANVAEYGLAGE"
10277..10717
/gene="AGR_C_1417"
10277..10717
/gene="AGR_C_1417"
/note="Ribbonuclease H (E.C.3.1.26.4) Mutant with His 62
Replaced By Pro (H62p)"
/codon_start=1
/transl_table=11
/product="AGR_C_1417p"
/protein_id="AAK86585.1"
/db_xref="GI:15155751"
/translation="MKHVDLFTDGCAGSGNPGQGVAVLRGTEKELSGCEADTTNN
RMELLAALNALKSPCEVDLYTDSAYVKGDTIKVIPEMKKGWMTADNKPVKNEVL
MQLLEAAOEHHKVTLLHWKAGHGPENERADELARKGMEPPKRR"
BASE COUNT 2021 a 3194 c 3242 g 2378 t
ORIGIN

alignment_scores:
Quality: 1872.50 Length: 548
Ratio: 4.018 Gaps: 5
Percent Similarity: 85.036 Percent Identity: 63.869

alignment_block:
us-09-712-768-2 x AE008010

```

```

Align seg 1/1 to: AE008010 from: 1 to: 10835
7 HISGLYHISASPHISHSISGLULYSGINGLYPHEPHEHTRARGTPHME 23
2031 CATTCGCATGATCATTCCTCAAGCCGGGCTTTTGGCCCGTGGTTCTT 2080
23 tserthrasnhsislyasphileglyleuleuleutyrleuvalaialagly 40
2081 GTCCACCAACCAACCAAGGACATCGGCACCTGTACCTGATCTTCGCCATCA 2130
40 aValIGlyPheIleSerValleuPheHtrValTyrMeAArgLeuLeu 56
2131 TGGCGGATATCATTCGGCGGCGGCTTTGGTCGTCATCGATGAGAGCTG 2180
57 MetAspProGlyValIGlnTyrMeCysLeuGlnGly...AlaArgLeuI 72
2181 CAGGAGCGCGGCATCCAG.....ATCTTCCACGGTCTGGCTTCATGCT 2224
72 eAlaAspAlaSerGlnThrCysThrAlaAsnGly.....HisLeuTrpA 87
2225 CTACGGCTTCGAGCGCATGCGCCATCGACGGGATAGACATGTTCA 2274
87 snValMetValThrTyrHisGlyIleLeuMetMetPhePheValIGlyLe 103
2275 ACGTGTTCACGACACGCGCGGATCATGATCATCTTCTCATGTGTCATG 2324
104 ProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetProLeuGlnIleG 120
2325 CCGGAGATGATCGGGGTTTGGCAACATGATGATCCGATCATGATGATCGG 2374
120 yAlaProAspMetAlaPheProAlaGmetAsnLeuSerPheTyrLeuP 137
2375 CGCGCTTCACATGCTTCCCGCGCCT.AMACATATTTCTTCTGCGTCA 2423
137 heIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAlaProGlyG 153
2424 TCGTTCGGCGCTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2473
154 AspGlyGlnLeuGlySerGlyValIGlyTyrValLeuTyrProProLeu 170
2474 GCGGCTGCTTATGCGCGTGGTGGCGGCGGATGATGATGATGCGCTGCTG 2523
170 TrhArg...GlnAlaGlyTyrSerMetAspLeuAlaIlePheAlaValH 186
2524 GACAGATGACATGCCCGCGCGCGGCGGATGATGATGATGATGATGATG 2573
186 IsteuSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIleHtrTh 202
2574 ACGTCGCGCGGATGATCTCGATCTCGCGCGCGCATCAACTTCATCACAC 2623
203 PheLeuAsnMetAlaGAlaProGlyMetThrLeuHisIlyValProLeuP 219
2624 ATCTGTAACATGCCCGCGCGCGGATGATGATGATGATGATGATGATG 2673
219 eSerTyrSerIlePheIleThrAlaTrpLeuIleLeuLeuAlaLeuPro 236
2674 TGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2723
236 alleuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThr 252
2724 TTTTGGAGGTGATCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2773
253 ThrPhePheAsnProAlaGlyGlyAspProIleLeuTyrGlnHisI 269
2774 GCGTCTTCTGCGCGGAGGCGGCGGATGATGATGATGATGATGATGATG 2823
269 eleuTrpPhePheGlyHisProGlyValTyrIleIleIleLeuProGlyP 286
2824 GTTCGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2873
286 heGlyIleIleSerHisValSerThrPheSerIlySerProValPhe 302

```



```

1502 GCTATACGCGCTGTCACCGCCGCTGATCATGATCTTCTTCATG 1453
102 GYILEPRLALeuphegIyIyphgIyAsnTYrLeuMetProLeuG1 118
1452 GTGATGCTGCGATGATGCGGGGTTGCGCAACTGTTGCTTCGATCAT 1403
118 nIlegIyAlaProAspMetAlaPheProArgMetAsnLeuSerPheT 135
1402 GATCGGCGCGCGGACATGCGCTTCCGCGGATGACAAACATCTGCTCT 1353
135 rPhePheIleAlaIyThrAlaMetGlyValAlaSerLeuPheAlaPro 151
1352 GGTGCTGATCGCGCGCTGATCTCTGCTGCTGCTGCTGCTGCTGCTG 1303
152 GYILEPRLALeuphegIyIyphgIyAsnTYrLeuMetProLeuG1 168
1302 GCGGCGCGCGCGCAC...GGCTTGGCGGGGTGGAGCATGATCTTCGCG 1256
168 oleuSerThr..ArgGluAlaGlyTyrSerMetAspLeuAlaIlePheA 184
1255 GCTGTGACCATGCGTACAAAGGTCGCGGCGCATGACATGCGCATCTGT 1206
184 IAlaAlaHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIle 200
1205 GGTTCACCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1156
201 ThrThrPheLeuAsnMetAlaGlyAlaIleThrMetLeuThrAspArgAsn 217
1155 ACCAGCATCTTCAATATGCGCGCGCGCATGACCTTGCACCGCATGCG 1106
217 oleuPheSerThrSerIlePheIleThrAlaTrpLeuIleLeuLeuAla 234
1105 GCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1056
234 eupProValIleuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsn 250
1055 TGGCGCGTCTGGCGCGCGCATGACCATGCTGCTGCTGCTGCTGCTGCT 1006
251 GYILEPRLALeuphegIyIyphgIyAsnTYrLeuMetProLeuG1 267
1005 GGCACCGCATCTTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 956
267 nHisIleLeuTrpPhePheGlyHisProGluValTYrIleIleLeuP 284
955 GCACCTGTTGCTGCTTTCGCGCGCACCGGAACTGATATCTGATCTGCG 906
284 roGlyPheGlyIleIleSerHisValSerThrPheSerLysPro 300
905 CGGCTTGGCATGATGACGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 856
301 ValPheGlyIyTrpLeuProMetValTYrAlaMetValAlaIleGlyVal 317
855 GTCCTTGGCATGATGACGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
317 uGlyPheValTYrAlaHisHisMetTYrThrValGlyMetSerLeuP 334
805 CGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 756
334 hrgInGlnSerTYrPheMetLeuAlaThrMetValIleAlaValProThr 350
755 ACCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 706
351 GYILEPRLALeuphegIyIyphgIyAsnTYrLeuMetProLeuG1 367
705 GCGCTGAGAGATCTTACCTGATGCGCAACATGCGCGCGCTGCTGCTGCT 656
367 uPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPheThr 384
655 CTTCAGAGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
384 aIGlyGlyValThrGlyIleValIleuAlaGlnAlaGlyLeuAspArgAla 400

```

```

605 TCGGCGCGCTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 556
401 TyrHisAspThrTYrTYrValValAlaHisPheHisTYrValMetSerIle 417
555 CTGACGACGACCTTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 506
417 uGlyAlaIlePheAlaIlePheAlaGlyIleTYrPheTYrMetProLysP 434
505 GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 456
434 hSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTrpThr 450
455 TGTGGGCGGTGAGTACAACGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 406
451 PhePheIleGlyAlaAsnValTYrPhePheProGlnHisPheLeuGlyArg 467
405 AAGTTCATGCGGTGACCTGATCTTCTTCCGCGACGCTTCTGCTGCTGCT 356
467 gGlnGlyMetProArgArgTYrIleAspTYrProGluAlaPheAlaLeuP 484
355 GCAAGGCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
484 rPAsnLysValSerSerTYrGlyAlaPheLeuAlaPheAlaSerPheLeu 500
305 GCACTACGCTGTCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
501 PhePheIleValIlePheValTYrThrLeuValAlaGlyArgArgGluTh 517
255 GGTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 209
517 rArgProAsnProTYrGlyGluPheAlaAspThrLeuGluTrpThrLeuP 534
208 CGAGCGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 159
534 rSerProProAlaHisThrPheGluThrLeuProLysArgSerAsp 550
158 CCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 109
551 TrpAspLysHis 554
108 GACGACGACGAC 97

seq. name: gb_AP03011
seq. documentation block:
LOCUS AP03011 346510 bp DNA BCT 15-MAY-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 18/21.
ACCESSION AP03011 BA000012
VERSION AP03011.2 GI:14026664
KEYWORDS
SOURCE
ORGANISM
Mesorhizobium loti (strain:MAFF303099) DNA.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Ido, S., Ishikawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsunoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpou, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
JOURNAL
MEDLINE
21082930
REFERENCE
2 (bases 1 to 346510)
AUTHORS
Kaneko, T.
TITLE
Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/~chizobase/
Tel: 81-436-52-3935 (ex. 2338), Fax: 81-436-52-3934)

```

```

COMMENT      On May 11, 2001 this sequence version replaced gi:11994986.
FEATURES
source
    1. 346510
        /organism="Mesorhizobium loti"
        /strain="MAFF303099"
        /db_xref="taxon:381"
        complement(242..1264)
        /gene="m117087"
        complement(242..1264)
        /gene="m117087"
        /note="similar to lipopolysaccharide core biosynthesis
        glycosyl transferase Lpse"
        /codon_start=1
        /transl_table=11
        /protein_id="BAB53261.1"
        /db_xref="GI:14026665"
        /translation="MAGLACAVRLVSDVHLPRILATADRGNAIDKRTONLRYLER
        REIYORQHDVYESDVNAADIVLIFTFOAKLEPLPSVLARSDRLVIGTSHREL
        ESEFREPGLATHRLARAVYNNRKLHEVAPLRIPIVHTPGVDVTFECRPEPOP
        YSGELRVAGMSLGNHGPARGFHDVIEPAVAAGARLTAIREOHNRHDEMIDFY
        RDLVYVCASSEGTNPCLPAACGIPLVTRVGNMPELVQDQDNGLFFDGTGELA
        DKIALLRDTPSLRSRMARMIETIRGMDVRVQAEVYAHMFTSLTAGAVPEVRAS
        DVGQY"
        complement(1228..2265)
        /gene="m117088"
        complement(1228..2265)
        /gene="m117088"
        /note="similar to lipopolysaccharide core biosynthesis
        glycosyl transferase Lpse"
        /codon_start=1
        /transl_table=11
        /protein_id="BAB53262.1"
        /db_xref="GI:14026666"
        /translation="MALAGEHDLQGVSTKGCALAMPSSLRLVISAENSCALIV
        ERPAMHPAYNEGLRNELVNTLTAHARENSPRVAVLITRNEERFVAGCENLEH
        OGVOYICDNOSTRTLEIAORYGAGLIGESIPIHNGWTCMOELLRRKEPRLQSLFA
        DMLHLDDEHLHPPTSSLSVAIAADALGPNATSSSEPTFPTREADHDNDPO
        HTLRTYFRFSPHCYRAKKQDGPETIANGSHKRRFBQVPLSEPRKMYIFL
        SEHAARKYGGRRFLSEVDGRGWGMPRLRPDIQLPDITQVTTATDDDEANP
        WLAMHLYAS"
        complement(2648..3787)
        /gene="m117089"
        complement(2648..3787)
        /gene="m117089"
        /note="similar to mandelate racemase"
        /codon_start=1
        /transl_table=11
        /protein_id="BAB53263.1"
        /db_xref="GI:14026667"
        /translation="MSLSHRITRFQFARBRVIGDSQVADVNAALIELVSGEYG
        LGFTQTFNPLPDQETESFEHEWPSLGNRAIVLHVNRPRGNGQRAYSLPEHE
        AVQVALMDLAKENGLPLHLVLRNRNRKAYASGLDFHLDDEAFVSLFSHAISIGS
        AFKIKVGRDFRDRLRLLELKCVPKNSVYMDIDPEAWTSKALTFLVAIRAGHDL
        LVMEDPLLRHHDGILRTLRHAHWTQINSEYLDLQGRLLLEHAADILNVGOYTD
        VMRGLLAELGIPISIGTNGFLEAGVMAVALPEVELEVSFQNFHDLVQPIETIDG
        YAVAPDRGPGIVLSEKARGEMSRRLANSELGAEPENRLPAK"
        3870..4556
        /gene="m1r7090"
        3870..4556
        /gene="m1r7090"
        /codon_start=1
        /transl_table=11
        /product="transcriptional regulator"
        /protein_id="BAB53264.1"
        /db_xref="GI:14026668"
        /translation="MAGKAKTSSPLVAARMSDIAYERILLEGFLDRVPAGAFISON
        ELSAIVDPVAPLRDALVLEAGILTIHPRSGIOFVRPGLTFRATYQORSIVEEAA
        VRVAFEGDEAVNNALEIRHSRLIRYERGLCKEQLDEBDALLEGELRGAVIAALNP
        LIDSTYRHMNYLRLRLDKRITTPIMVTKLEHDLDEACGRNADAEAKQLAHFQ
        AAQGRSLGLV
        4684..5922
        /gene="m1r7091"
        4684..5922
        /gene="m1r7091"

gene
CDS
    /gene="m1r7091"
    /codon_start=1
    /transl_table=11
    /product="sugar ABC transporter, periplasmic binding
    protein"
    /protein_id="BAB53265.1"
    /db_xref="GI:14026669"
    /translation="MQLNRQFLMSGNALAMGHINGARASAISTWHIFTQSENAS
    LKATIGLFOANADVVTQENTIPNSEYMAKVSVAVLGGRPTAVYIAERFADLTAME
    GLIDLTRVNGMGKAFENRWGTGLSSNGAIYAPAFVDMVYRADYFEBAGLAD
    PPRNFDEFSLACRKLTDPSKRYAFGRAGAGGFKYVIDMEAGSPYKDCQPAIDR
    AAAIEATFEYSDLEFKKAYVPSAPNDSTYQIIEGFRGTQTAWHHGTSLNEISAM
    KQKQFSTAPMPAGPKAHVARVAYAGNGIKEDNIEAARMVSPMGCTGAALTILECT
    GYPPSAALADDPRTGNPIYHAAVETLEGRPLPNFVGAAGSENVVPAQVAVLTG
    OSPEQAVPMIDGLEALR"
    5979..6932
    /gene="m1r7093"
    5979..6932
    /gene="m1r7093"
    /codon_start=1
    /transl_table=11
    /product="ABC transporter, permease protein"
    /protein_id="BAB53266.1"
    /db_xref="GI:14026670"
    /translation="MLSHRGSMTSPSSGYPRLALGELSETRYMYLILPLSLMILL
    VIAYPTLYGMAISFREMRRLRPGICGVGFKHYAAMWSPVFMILKNTVWVTAAY
    LELALGLIATALNRDLPCKTIFAVLILPYELPNVYAGHMAALLIDPRLQINDLV
    RAGVQLRTAMPADPNTALAAVYVAVHNGFPFALLLAGKSLPEDITKAAAYDGA
    GYNAOLRLITPLMKVITAAVILRVITGLVNSDILLITLGGPDQATQVLSLVAFT
    AKENFGIASALSVVMEVILMFATVYIRLSIRIKGDD"
    6925..7770
    /gene="m1r7094"
    6925..7770
    /gene="m1r7094"
    /codon_start=1
    /transl_table=11
    /product="ABC transporter, permease protein"
    /protein_id="BAB53267.1"
    /db_xref="GI:14026671"
    /translation="MIESKRRRRVIANITVPLVLTISFCIGPMIHWELTSKLEAD
    IYQIMQYIPRVPTPNVYAITQSGEPRLTANSLVYPTLVTICVATGTLAAVFSR
    FARGRTQLMLGTLIVRMFPAYMMIPLTVVRGVLVDSRGLAATSEFLPLFV
    MLKGFDSAPKELESARIDGSRUGAMVRYPLRNGLVASSVFIAMANEFLFA
    LMLTGGGSRTPVGLQLVGFEQLPMGVLAGGILSTILYVLFATVQRTVQGLTA
    GAIK"
    7781..8860
    /gene="m1r7095"
    7781..8860
    /gene="m1r7095"
    /codon_start=1
    /transl_table=11
    /product="sugar ABC transporter, ATP-binding protein"
    /protein_id="BAB53268.1"
    /db_xref="GI:14026672"
    /translation="MATLSIKDVRSYSGVOYLHGVDIELEDGFLVLLGSGGCKST
    LNLIIAGLDLDTSGDILIGRSVVDLAKDRMIAYFOSYALPTMSVERNIGGLEEM
    RGVDADQRRKRAVDAARLLOISHLDRPANLSGROVRVAMGRAIVADPELPEFDP
    LSLNDAKLREVMERTKRLHERLGTISYVTHDOIEMATLAKTAVMGHIIQOPADP
    RTIYERPAVAVASFISGPMNPLDGRLYSATGYRPAASGVVALQGRKTVYSER
    VILGRPELKYACGDSEASVYTDVYERTGPEHWTVQVQDOSTIARLPREFGRN
    EPIFLAVDPASVNLDFPGSEHRIDI"
    8864..9082
    /gene="msr7097"
    8864..9082
    /gene="msr7097"
    /note="unknown protein"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB53269.1"
    /db_xref="GI:14026673"
    /translation="MDAKALLAARGFGHPRHSRRHRCRDRKRLRLQRLMAAVALAY
    AGCYRLYEINAMIPMGRTAARART"
    9079..10692

```

```

CDS
/gene="mlr7098"
9079..10692
/gene="mlr7098"
/codon_start=1
/transl_table=11
/product="GMC-type oxidoreductase"
/protein_id="BAB53270.1"
/db_xref="GI:14026674"
/translacion="MSETSPFGDYDIYVAGSAGCYLANRLSNDRNRVLLIAGSS
DRHMDVPIGLYFCMGNPRDWMKTEPEAGLNLNTPRGKVLGGSSVMGMITM
RQADADYGRKQAGNMGWSDVDLYPRKSEDDHGGKSAHSGGGEWRVERORLSWT
LDARDAEELGELPRTGDFNTGMEGSGYEENVRGLRVNNTSKAFIRGIAAQNLRV
VTGARETELEEDGRVYGLLEPGDRHLRVSAGOTIIAAGINSPIILEISGVRPL
LSISGVVRHALLPGVGNLIDHLOIRTFVAVANATILNERYHNLISRAAMGVEALRK
SGPLSMASPOLGIFARSNDARLPTDELYHNQPLSTDRGRLHFRPAVTVSCNLRPE
SRGVTHMASSDPREARKIRREYLTSTAGDSVAASLHNVSLMKARVAFAPREMLP
CTOYDSEDLIRKAGDITGLTFHVGCKMGSDSMAVVDRLRVHGLAGLVVDASTM
PTIVSGNTNSPVVIAEKAAEMILLDPR"
10929..11801
/gene="mlr7099"
10929..11801
/gene="mlr7099"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="BAB53271.1"

alignment_scores:
Quality: 1846.50      Length: 555
Ratio: 3.937          Gaps: 8
Percent Similarity: 84.505      Percent Identity: 61.441

alignment_block:
US-09-712-768-2 x AP003011 ..

Align seg 1/1 to: AP003011 from: 1 to: 346510

1 MetAlaSPaLaLaLaLeHs...GLYHLSAsPHisHISgluLyscIngl 16
|||||
337162 ATGGCAGACCTGCTGACGTCACGACGGCCAGACCAAGCCTTATCATG 337211
16 yPhePheThArGTTrPheMetSerThraSnHLSysAsPILeul 33
| : : : : :
337212 C...TGSGTGGCGGTGCTATTGACCAACCAAGACATCGGCACGC 337258
33 eUrYrLeuValAlaLaGlyValAlaGlyPheLeSerValLeuPheThr 49
|||||
337259 TCCTACCTGATCTTCGCGATGCGCGCATGCAATCGGCGCGCTGTG 337308
50 ValYrMetArGLeuGluLeuMetAsPProGlyValGlnTrMetCysLe 66
|||||
337309 GTCCGCATCCGCATGAGTTCGACGACCGCGCATCCAG...ANCTT 337352
16 uGlUGly...AlaArGLeuIleAlaSPaLaSerGlnThCysThraLa 82
| : : : : :
337353 CAGCGGTCTGCGCAGATGTCTACGCGCATGACGTCGACGCGCATTTG 337402
82 snGly.....HisLeuTrAsnValMetValThrYrHisGlyIleLeu 96
|||||
337403 ATGGCGGCGACAGCATGTACATGCTTCCGACCCGCGCATGCGCTATC 337452
97 MetMetPhePheValGlyIleProAlaLeuPheGlyIlePheGlyAsnTy 113
|||||
337453 ATGATCTTCTTCATGTCATCCGCGGTGATCGCGGCTTCGCAACTG 337502
113 rLeuMetProLeuGlnIleGlyAlaProAsPMeAlaPheProArgMeTa 130
| : : : : :
337503 GATGGTGGCATGATGATGCGCGCATGATGCTCCCGGCAATGA 337552
130 snAsnLeuSerPheTrPhePheHeLaGlyThrAlaMetGlyValAla 146
|||||
337553 ACAACATCTCTCTTGGCTGCTGCGCGCGCTTCATCTGCTGCTACCC 337602

```

```

147 SerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTr 163
|||||
337603 TCAGATCTTCGTCGCGCAGTGGCCCGGCGCCTTCGCGCGCGCGGCTG 337652
163 pValLeuTyTrProProLeuSerThraG...GlnAlaGlyTySerMeTa 179
| : : : : :
337653 GACGCTCATTCGCGCGCTGCTGACCTCGGTCAGCCGCGGACCGCATGG 337702
179 sPLeuAlaIlePheAlaValHisLeuSerGlyAlaSerIleMeGly 195
|||||
337703 ATCTGGCGCATCTGTCGATCCACATTCGCGGTGCTGCTGATCTGCTGGC 337752
196 AlaIleAsnMetIleThrThrPheLeuAsnMetArGAlaProGlyMetTr 212
| : : : : :
337753 GCCATCACTTCATCACACCATCATTCACATGCGCGCTCCGCGCATGAC 337802
212 rLeuHisLysValProLeuPheSerTrpSerIlePheIleThrAlaTrpL 229
|||||
337803 GCTGCACAAGATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337852
229 euIleLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeu 245
| : : : : :
337853 TGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337902
246 ThrAsParGAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAs 262
| : : : : :
337903 ACCGACCGCAATTCGCGCACCATCTTCGCGCGCGCGCGCGCGCGCA 337952
262 pProIleLeuTyTrGlnHisIleLeuTrPhePheGlyHisProGlyValT 279
| : : : : :
337953 TCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338002
279 yrIleIleIleLeuProGlyPheGlyIleIleSerHisValAserThr 295
| : : : : :
338003 ACATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338052
296 PheSerLysLysProValPheGlyTyTrLeuProMetValTyAlaMetVa 312
| : : : : :
338053 TTCTCGAAGAACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338102
312 lAlaIleGlyValLeuGlyPheValValTrAlaHisHisMetTyTrThv 329
| : : : : :
338103 CCGCATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338152
329 aGlyMetSerLeuThrGlnGlnSerTyTrPheMetLeuAlaThrMetVal 345
| : : : : :
338153 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338202
346 lIleAlaValProThrGlyIleLysIlePheSerTrpIleAlaThrMetTr 362
| : : : : :
338203 ATCGCGGTCCGCGACGCGGCGTGAAGATCTTCCTGATCGCATGCGATGTG 338252
362 pGlyLysSerValGluPheLysSerProMetLeuTrAlaPheGlyPheM 379
| : : : : :
338253 GGGCGGCTGATCTGCTTCAAGACGCCGATGCTGTGGCGCGCGCTTCA 338302
379 eTrPheLeuPheThrValGlyGlyValThrGlyIleValLeuAlaGlnAla 395
| : : : : :
338303 TCCTCTGTTACCATCGATGCGGTGCGTCACCGGCTTCAGCTGCGCATGCG 338352
396 gLyLeuAsParGAlaTyTrHisAsPTrTyTyValValAlaHisPheH 412
| : : : : :
338353 GGGCTGACCGCGCTGCGATGACACCTATTTGATGATGCGCCACTTCA 338402
412 sTyValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyTr 429
| : : : : :
338403 CTACGTGCTGCTGCTGCGCGGCTGTTCGCGCATCTTCGCGGCTGCTACT 338452
429 heTyMetProLysPheSerGlyAlaAlaPheProGluTrAlaAlaLys 445
| : : : : :
338453 ACTGCTGCCCAAGATGACCGCGCTACATGATTTGCGCCGCTCATGCGCAAC 338502
446 LeuHisPheTrPhePheHeIleGlyAlaAsnValThrPhePheProG 462

```



```

1879 CATACAG.....GTTTTGATCCGCG. 1900
77 InthrcysThrAlaAsnGlyHisLeuTrpAsnValMetValThrGlyHis 93
1901 .....CATACCTACACGCTGTTTCCGACCAATCCAC 1930
94 GlyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGlyGlyPh 110
1931 GGCTCATATCATATCTTCTTCATGGTATGCTCGCATGATCGCGCGCTT 1980
110 eGlyAsnTyrlLeuMetProLeuGlnIleGlyAlaProAspMetAlaPheP 127
1981 CGGCACACTGTTGCTGCGCTGATGATCGCGCGCGCCGACATGCGCTCC 2030
127 roArgMetAsnLeuSerPheTrpLeuPheAlaIleGlyThrAlaMet 143
2031 CGGCATGTGAACAACATCTGTTCTGGCTGCTGCGCGCGCTTGGCGTG 2080
144 GlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySerS 160
2081 CTGCTATATCTCTCTTCATGGAGGTGAGCCGCGCGCGACCGCGAGCGG 2130
160 yValGlyTrpValLeuTyProProLeuSerThrArg...GluAlaGlyT 176
2131 CACGGGCTGGACGCTGATGCGCGCTGACGACGCGTCAATCCGCGGTC 2180
176 ySerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAlaSer 192
2181 CGCGGGGCGCTCACATCTGTCATCCACCTTCGCGCGCTCTCTCG 2230
193 IleMetGlyAlaIleAsnMetIleThrPheLeuAsnMetArgAlaPr 209
2231 ATCCTCGGGCGCATCACTTCATACACGATCTTCACATGCGCGCGC 2280
209 OGlyMetThrLeuHisLysValProLeuPheSerTrpSerIlePheIleT 226
2281 TGGCATGACCATGCACAGATGCGCGCTGTCGTCGTCGATCTGTGTA 2330
226 hrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAlaIleThr 242
2331 CGGCTTCTGCTGCTGCTGCGCGCTGCGCTGCTGCGCGCGCGATCAGC 2380
243 MetLeuLeuThrAspArgAsnPheGlyThrThrPhePheAsnProAla 259
2381 ATGCTGTGACGACCGGACTTCGCGACACGCTTCTCGCGCGCATG 2430
259 yGlyGlyAspProIleLeuTyrlGlnHisIleLeuTrpPhePheGlyHisP 276
2431 CGCGCGGCGATCCGCTGCTGTCAGCATGCTGCTGCTGCTGCTGCTG 2480
276 roGlyValTrpIleIleLeuProGlyPheGlyIleIleSerHisVal 292
2481 CGGAGGTGATCATCTGATTCGCGCGCTTCGCGATGATCACCACATG 2530
293 ValSerThrPheSerLysProValPheGlyTyrlLeuProMetValTy 309
2531 ATCTCGACGCTTTCGAAGAAGCCGATCTTCGATATCTGGGAGTGGCTA 2580
309 rAlaMetValAlaIleGlyValLeuGlyPheValValTrpAlaHisHisM 326
2581 CGGCATGTCGTCGATCGCGCATCGCTTGTCTGCTGCGCGCCACACA 2630
326 eTyrlTrpValGlyMetSerLeuThrGlnIleSerTyrlPheMetLeuAla 342
2631 TGTACAGGTGGGATGCTGCTCTCGCGCGACGCTACTTCTGTCGCGGCG 2680
343 ThrMetValIleAlaValProThrGlyIleLysIlePheSerTrpIleAl 359
2681 ACCATGTGTATCGCGGTGCGCGAGCGGTGAAGTGTTCGTGATGATCCG 2730
359 aThrMetTrpGlyGlySerValGluPheLysSerProMetLeuTrpAlaP 376

```

```

2731 CACGATGTGGGCGGCTCGATCGAGTTCAGAGCCCGCATGCTGTTCCGA 2780
376 heGlyPheMetPheLeuPheThrValGlyGlyValThrGlyIleValLeu 392
2781 TCGGCTTATCTCTCTGTTCCACGTCGCGCGCGCTGACCGCGCTGCTG 2830
393 AlaGlnAlaGlyLeuAspArgAlaIleTyrlHisAspThrTyrlValAla 409
2831 GCCATGCGGGGCGTTCATGCGCGTGTGATGACAGACACTACATGATCCG 2880
409 aHisPheHisTyrlValMetSerLeuGlyAlaIlePheAlaIlePheAla 426
2881 GCACCTTCACACTACGATGATGCGCATGCGCGCTGTCGCGATCTTCGCG 2930
426 LyIleTyrlPheTyrlMetProLysPheSerGlyArgAlaPheProGlu 442
2931 GCTGTTACTACTGTTCTCCAGAGATGTCGGCTACATGTACACGACAGCC 2980
443 AlaAlaLysLeuHisPheTrpThrPhePheIleGlyAlaAsnValThrPh 459
2981 ATCGGCAAGCTGCACCTCTCGGATGATGATTCATCGCGCGCACATCTGT 3030
459 ePheProGlnHisPheLeuGlyArgGlnGlyMetProArgArgTyrlIleA 476
3031 CTTCGCCGACGACACTTCCTCGCTGTCAGGCGCATATGCGGCGCATGTG 3080
476 sPTyProGluAlaPheAlaLeuTrpAsnLysValSerSerTyrlGlyAla 492
3081 ACTATCCGATGCTCTTCGCGGCTGGAACGATCTCTCTACGCGCGCC 3130
493 pheLeuAlaPheAlaSerPheLeuPheIleValIlePheValTyrlTh 509
3131 TTCATACAGGCGC.....TTCGCGCTCATCATCTCTCTACGCG 3168
509 rLeuVal.....AlaGlyArgArgGluThrArgProAsnProTrpG 523
3169 CTTGCTCGACGCTTCGCGCGCACAGACAGCGGCTGACATCCGTGGG 3218
523 LyGluPheAlaAspThrLeuGluTrpThrLeuProSerProProAla 539
3219 GTGAGAGCGCGGACGACGCTGGAATGACGCTGCTCACCAGCTCTCGTTC 3268
540 HisThrPheGluThrLeuProLys 547
3269 CACCACTTCTCCACCTGCGGAG 3292

seq_name: gb_ba:BCOXAG

seq_documentation_block:
LOCUS      BCOXAG               1830 bp      DNA          BCT          12-SEP-1993
DEFINITION B.japonicum coxa gene for cytochrome aa(3) subunit I (BC 1.9.3.1).
ACCESSION  X54318
VERSION    X54318.1 GI:39505
KEYWORDS   coxa gene; cytochrome aa(3) subunit I.
SOURCE     Bradyrhizobium japonicum.
ORGANISM   Bradyrhizobium japonicum
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Bradyrhizobium group; Bradyrhizobium.
REFERENCE  1 (bases 1 to 1830)
AUTHORS    Gabel,C.
TITLES     Direct Submission
            Submitted (07-AUG-1990) Gabel C., Johns Hopkins University, Dept.
            of Biology, Charles and 34th Sts., Baltimore, MD 21218, USA
            2 (bases 1 to 1830)
AUTHORS    Gabel,C. and Meier,R.J.
TITLES     Nucleotide sequence of the coxa gene encoding subunit I of
            cytochrome aa3 of Bradyrhizobium japonicum
            Nucleic Acids Res. 18 (20), 6143 (1990)
JOURNAL    MEDLINE
FEATURES   91045095
            Location/Qualifiers
            1..1830
            /organism="Bradyrhizobium japonicum"
            /strain="LO"

```

```

RBS      /db_xref="taxon:375"
167..170 /note="ribosome binding site (pct.)"
178..1803 /note="cytochrome aa(3) subunit I"
CDS      /codon_start=1
         /transl_table=11
         /protein_id="CA38216.1"
         /db_xref="GI:39506"
         /db_xref="SMISS-PROT:P31833"
         /translation="MARSAAHGDHGDHGHHAHPGMRKRYVSTNKHDKITWYLI
         FAVIAYGTAAMSIATRAELMTPGVOIFHEHTYVNTSHGLIMIFEMVPMATIGF
         GWNEVPLMGADPMAPRNNISFWLPSFGLLMTSTFEGEPGANGVAGMTWVP
         LSSSGHPGPAVDFAILSLHLAGASILGAINFTITFMKAPGMTLHKMPLFVNSILV
         TVFLLSLPVLGATITMLTDNRNGTTFEADGDDPYLFOHLMFPGHPRVYLLIL
         PCGMASIVSTFESKRPVGYLGAMAYAMVIGIGVYVAHHMTYVSSATQAYEVA
         ATMVATAPGVKIFSWIATMMGSGIIFRAPMTWAGELFLPTVGSGTVYLANAGVDR
         VLOETTYVVAHRYVSLGAVVAIFRAGTYMPKMGTYGNTETLAKHFWTFEIGNL
         VEPQHFELSGMPRKYVDYPDAFGWNLVSSVSGVISGFLIFLYCVIDAFKKRP
         ADDNPWAGATTLLEMTLPSPPFHQEVLPRVO"
BASE COUNT      282 a      606 c      566 g      376 t
ORIGIN
alignment_scores:
      Quality: 1829.00      Length: 545
      Ratio: 4.073      Gaps: 4
Percent Similarity: 82.385      Percent Identity: 61.835
alignment_block:
US-09-712-768-2 x BJCOXAG ..
Align seg 1/1 to: BJCOXAG from: 1 to: 1830
7 HisGlyHisAspHisSHISGLuLysGlnGlyPhePheThrArgTrpPhe 23
  |||||.....:|||||:  :|||:
220 CACGGACATGACGAGACACGCCATCCAGCGGATGCGGACGCTACGCTA 269
  tSerThrAsnHisLysAspIleGlyLeuLeuTyrlLeuValAlaIaGly 40
  |||||.....:|||||:  :|||:
270 TTCGACCAACACACAAGCATCGGACGATGCTACCTGATCTCGCGGTG 319
  40 aIValGlyPheIleSerValLeuPheThrValTyrlMetArgLeuGlu 56
  :|||:|||||:  :|||:
320 TCGCGGCGATGATGCGCGCATGTCGATCGGATCGCGGACGCTG 369
  57 MetAspProGlyValGlnTyrlMetCysLeuGluGlyAlaArgLeuIle 73
  |||  |||||.....:
370 ATGATATCGGCGGCTGCAGATCTTC..... 393
  73 aAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMet 90
  :|||  |||  :|||:
394 .....  CACGACGACGACCTACACAGCTTCG 421
  90 aThrTrpHisGlyIleLeuMetMetPhePheValGlyIleProAlaLeu 106
  |||||  |||||.....:|||||:  :|||:
422 TGACCTCCACGCGCTGATCATGATCTTCATGATGATGCGCGCATG 471
  107 PheGlyIlePheGlyAsnTyrlLeuMetProLeuGlnIleGlyAlaPro 123
  :|||:|||||:  :|||:
472 ATCGCGGCGCTTCGGCACTGCTTCGCTGATGATGCGCGCGCGCA 521
  123 PheAlaPheProArgMetAsnAsnLeuSerPheTrpLeuPheIleAla 140
  |||||  |||||.....:|||||:  :|||:
522 CATGGCGCTTCGCGCATGAACAACATCTGCTGCTGCGCGCGCT 571
  140 LysThrAlaMetGlyValAlaSerLeuPheAlaProGlyIleAspGly 156
  :|||:.....:  |||  |||:
572 CATTGCGGCGCTGCTGATGCTGCTGCGAGCGCGCGCGCGCG 621
  157 LeuGlySerGlyValGlyTrpValLeuTyrlProLeuSerThrArg... 172
  |||  |||||.....:|||||:  :|||:
622 AACGGCGCTGCGCGCGCTGACCATGTACGTGCGCTGCGAGCTCGG 671
173 .GluAlaGlyTyrlSerMetAspLeuAlaIlePheAlaValHisLeuSer 189
  |||||.....:|||||:  :|||:
672 CCATCCCGCGCGCGCGAGTCTGCGATCTGCTGCTGCTGCGCG 721
  189 LysAspSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsn 205
  |||||.....:|||||:  :|||:
722 GTGCGCTGCTGATCTGCGCGCATCTGCTGCTGCTGCTGCTGCTG 771
  206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrp 222
  |||||.....:|||||:  :|||:
772 ATCGCGCGCGCGCGCATGACCCCTGCAAGAGATGCCCTGCTGCTG 821
  222 LysPheIleThrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAla 239
  |||||.....:|||||:  :|||:
822 GATCGCGTACCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCG 871
  822 ArgCysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 871
  239 LysAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhe 255
  |||||.....:|||||:  :|||:
872 GTGCGATCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 921
  256 AsnProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 921
  |||||.....:|||||:  :|||:
922 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 971
  272 ePheGlyHisProGlyValTyrlIleIleLeuProGlyPheGlyIle 289
  |||||.....:|||||:  :|||:
972 CTTCGCGCATCCGAGGTGATGCTGCTGCTGCTGCTGCTGCTGCTG 1021
  289 LysSerHisValValSerThrPheSerLysLysProValPheGlyTyrl 305
  |||||.....:|||||:  :|||:
1022 TCAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1071
  306 ProMetValTyrlAlaMetValAlaIleGlyValLeuGlyPheValTr 322
  |||||.....:|||||:  :|||:
1072 GGCATGCGCTACGCGATGCTGCGATGCGCGCGCGCGCGCGCTGCT 1121
  322 PAlaHisIleMetTyrlThrValGlyMetSerLeuThrGlnIleSerTy 339
  |||||.....:|||||:  :|||:
1122 GGGCGCACCATGATGACGCTGCGCATGCTGCGCGCGCGCGCGCT 1171
  339 hemeLeuAlaThrMetValIleAlaValProThrGlyTyrlLeuIle 355
  |||||.....:|||||:  :|||:
1172 TCGTCCGCGCGCGAGATGATGCGCGCTGCGCGCGCGCGCGCTG 1221
  356 SerTrpIleAlaThrMetTrpGlyGlySerValGlyIlePheLysSer 372
  |||||.....:|||||:  :|||:
1222 TCGTGAATGCGCGCATGCTGCGCGCGCGCGCGCGCGCGCGCT 1271
  372 LysThrAlaPheGlyPheMetPheLeuPheThrValGlyIleValThr 389
  |||||.....:|||||:  :|||:
1272 GATCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1321
  389 LysIleValLeuAlaGlnAlaGlyLeuAspArgAlaTyrlHisAspThr 405
  |||||.....:|||||:  :|||:
1322 GCGTCTGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCTGCT 1371
  406 TyrlValAlaHisPheHisTyrlValMetSerLeuGlyAlaIlePhe 422
  |||||.....:|||||:  :|||:
1372 TACGCTGCTGCGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1421
  422 aIlePheAlaGlyIleTyrlPheTyrlMetProLysPheSerGlyAla 439
  |||||.....:|||||:  :|||:
1422 GATCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1471
  439 hemeProGlyTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGly 455
  :|||  |||  |||||.....:|||||:  :|||:
1472 ACAGACGACGCGCTGCGAGCGCGCATCTGCTGCTGCTGCTGCTG 1521
  456 AsnValThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetPro 472
  |||||.....:|||||:  :|||:
1522 AACCTGCTGCTTCTGCGCGAGCATCTGCTGCTGCTGCTGCTGCTG 1571

```

```

472 gAgyrYrIleAspYrProGluAlaPheAlaLeuTrpaSnlYsValSerS 489
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1572 CCGCTACGTCGACTATCCGACGCGTCCGCGGCTGCAACGTCGTCTCT 1621
489 eRTYrYrIleAlaPheLeuAlaPheAlaSerPheLeuPheIleValIle 505
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1622 CGGTCCGCTTCACATTCCTCCGCC.....TTCCGCGCTCGATTC 1659
506 PheValYrTrhLeuValAlaGlyArgGluThrArgPro..... 519
1660 TTCTCTATTCGCTGATCGATCGCTTGGAAAAAAGATCCCGCTCGCA 1709
520 .AsnProTgIyGluPheAlaAspThrLeuGluTrpThrLeuProSerP 536
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1710 CAATCCGTCGGCGCGCGCGACGCTGGAGTGAGCGCTCCCTCC 1759
536 roProProAlaHisThrPheGluThrLeuProlys 547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1760 CGCGCGCTTCACATTCGATTCGAAAGTCTGCCCGCC 1794

seq_name: gb_ba:BJCOXA

seq_documentation_block:
LOCUS BJCOXA 2000 bp DNA BCT 07-NOV-1991
DEFINITION B.japonicum cox a gene for subunit I of cytochrome c oxidase
(EC_number=1.9.3.1).
ACCESSION X54800
VERSION X54800.1 GI:39502
KEYWORDS cytochrome c oxidase.
SOURCE Bradyrhizobium japonicum.
ORGANISM Bradyrhizobium japonicum
Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
Bradyrhizobium group: Bradyrhizobium.
1 (bases 1 to 2000)
REFERENCE
AUTHORS Bott M.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-1990) Bott M., Eidnoessische Technische
Hochschule, Mikrobiologisches Institut, Schmiedelbergstrasse 7,
CH-8092 Zuerich, Switzerland
2 (bases 1 to 2000)
REFERENCE
AUTHORS Bott M., Bolliger M. and Hennecke H.
TITLE Genetic analysis of the cytochrome c-a3 branch of the
Bradyrhizobium japonicum respiratory chain
JOURNAL Mol. Microbiol. 4 (12), 2147-2157 (1990)
MEDLINE 91211625
FEATURES
source
1..2000
Location/Qualifiers
/organism="Bradyrhizobium japonicum"
/strain="USDA311b110spc4"
/db_xref="taxon:375"
/clone="PRJ3404"
75..1700
/gene="cox a"
75..1700
/gene="cox a"
/codon_start=1
/transl_table=11
/product="cytochrome-c oxidase"
/protein_id="CA38570.1"
/db_xref="GI:39503"
/db_xref="SWISS-PROT:P31833"
/translacion="MATSAAAGHDHAKODGHDEHAFPTGMRVYVSTNKHDDIGTMYLI
FNYAVIGAAVMSIATRAELMYPGVOIFHEPTHYNVFVSHGLIMFPMVPMIAGF
GNMFVLMIGAPPMAPRRNNNTSPMLPASGILLMSTFVEBPGANGVAGAMTWVP
LSSGHPGPAVDAILSLHAGASSILGAINFTTTFNKRAKGTILHKRPILVMSITV
TYFLLSLPVLALGATITMLTDNFGTTFADGGEDPVLFOHLEFTEHPEYIIL
PGRGMSIQIYVTSRKPVFGYGLMAYANAIGIGLVVANAHHYTVGMSATQAYFA
ATWIAVPTGVKIFSVIATMGSGIEFRAMIAVGFILFTVGATGVVLNACGVR
VLOETYYVAHHFHVLSLGVFAIFAGWYVPMKGYMNETLIAAHFVMEFIGNLV
VFPOHFLGICSGMRVYVDPDAFAGMNVSSGYSISGVLIFLYCIDAFKAKVP
AGDNPWGAGATITLMTLPSPPPHOFVLPVVO"
75..1697
/mat_peptide
/gene="cox a"

```

```

CDS
/product="cytochrome-c oxidase"
1854..>2000
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CA38571.1"
/db_xref="GI:39504"
/db_xref="SPTRMBL:099253"
/translacion="MSLVITFALVGNAMPGHFHPVLATISLLCTAVGASGALNMA
LESDI"
1854..>2000
mat_peptide
BASE COUNT 304 a 669 c 614 g 413 t
ORIGIN

alignment_scores:
Quality: 1829.00 Length: 545
Ratio: 4.073 Gaps: 4
Percent Similarity: 82.385 Percent Identity: 61.835

alignment_block:
US-09-712-768-2 x BJCOXA ..
Align seg 1/1 to: BJCOXA from: 1 to: 2000
7 HISGLYHISASPHISGLULYSGINGLYPHEPETHRARGTRPHEME 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 CACGACATGACGACGACGACGACGACGACGACGACGACGACGACGACG 166
23 tserThrasnHisLysAspIleGlyLeuLeuTrpLeuValAlaIleGly 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 TTCGACCAACACAAAGACATCGCGACGATGACCTGCTTCGCGGTCA 216
40 aValGlyPheIleSerValLeuPheThrValTrpMetArgLeuGluLeu 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 TCCTCCGCTCATCGGTGCGCGGATGATCGATCGATCGATCGATCGATCG 266
57 MetAspProGlyValGlnTrpMetCysLeuGluGlyAlaArgLeuIleAl 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 ATGTATCCGCGCGCTGACATCTTC..... 290
73 asPalaSerGlnThrCysThrAlaAsnGlyHisLeuTrpaSnlYsVal 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 .....CACGACGACGACGACGACGACGACGACGACGACGACGACGACG 318
90 alThrTrpHisGlyIleLeuMetPhePheValGlyLeuProAlaLeu 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 TGACCTCCGACGCGCTGATCATGATCTTCTCATGGTCATGCCGCGGATG 368
107 PheGlyGlyPheGlyAsnTrpLeuMetProLeuGlnIleGlyAlaProAs 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 ATGGGCGCGCTTCGCAACTGCTTCGCGCGCTGATGATGCGCGCGCGCA 418
123 pMetAlaPheProArgMetAsnAsnLeuSerPheTrpPheIleAlaIle 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 CATTGCGCTCCCGCGCATGACATCTGCTGCTGCTGCTGCTGCTGCTGCTG 468
140 IyThrAlaMetGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGln 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 CTTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
157 LeuGlySerGlyValGlyTrpValLeuTrpProLeuSerThrArg.. 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 AACGCGCTCGCGCGCGCGCTGACCATGTACGCTCGCGCTCGCGCTCGG 568
173 .GluAlaGlyTrpSerMetAspLeuAlaIlePheAlaValHisLeuSer 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
569 CCAATCCGCGCGCGCGCATGCTTCGCGATCTGCTGCTGCTGCTGCTGCTG 618
189 IyAlaSerSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsn 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
619 GTGCTGCTGATCTGCGCGCATCAACTTCAATCAACGACGATCTTCAAC 668

```

```

206 MetArg1aProGlyMetThrLeuHisValProLeuPheSerTrpSe 222
|||||
669 ATGGCGCGCGCGGCGATGACCCCTGCACAAAGATCGCGTTCGTGGTGC 718
222 rIllePheIleThrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAla 239
|||||
719 GATCTGCTGACCGCTTCTCTGCTGCTGTGTGCTGCTGCTGCTGCTGCG 768
239 LysAlaIleThrMetLeuLeuThrAspArgAsnPhelGlyThrThrPhePhe 255
|||||
769 GTGGGATCACCATGCTGCTGCACGACCGCAATTTCGGACAGACTTCTTC 818
256 AsnProAlaGlyGlyAspProIleLeuTrpGlnHisIleLeuTrpPh 272
|||||
819 CGCGCCGACGCGCGCGGCGACCGCTGCTGCTGCAGCACCTGTCTGTT 868
272 ePheGlyHisProGlyValTyrIleIleIleLeuProGlyPheGlyIle 289
|||||
869 CTTGCGGCATCCGAGGTATCATCTGATCTCTCCCGGCTTCCGCGCATGA 918
289 IeSerHisValIleSerThrPheSerLysLysProValPheGlyTyrLeu 305
|||||
919 TCAGCCGATCGTCTCCACCTCTCGCGCAAGCCGTTGCGCTATCTC 968
306 ProMetValTyrAlaMetValAlaIleGlyValLeuGlyPheValValTr 322
|||||
969 GGCATGGCTTACGCGATGTCGCGATCGCGGCGATCGCTTCTGTGGTGTG 1018
322 PALAHSHISMetLysThrValGlyMetSerLeuThrGlnIleSerTrp 339
|||||
1019 GCGGCACACATGATGACGCGTGGCATGTGTCGCGGACCGACCGCTATT 1068
339 hemeLeuAlaThrMetValIleAlaValProThrGlyIleLysIlePhe 355
|||||
1069 TCGTCGGCGGACGATGTCATCGCTTCGACCGCGCGGATGATCTTC 1118
356 SerTrpIleAlaThrMetTrpGlyLysValGlyPheSerProLeu 372
|||||
1119 TCGTGGATCGCCACGATGTGGCGGCTGATGATTCGCGCGCGCAT 1168
372 tLeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyValValThrG 389
|||||
1169 GATCTGCGCGGCTGCTTATCTTCTGTCACGCGTGGTGGCTGCACCG 1218
389 LysIleValLeuAlaGlnAlaGlyLeuAspArgAlaThrHisAspTrpTr 405
|||||
1219 GCGTCTGCTGGGACGCGCGGCTGACCGCTGCTCCAGGAGACCTAC 1268
406 TyrValValAlaHisPheHisIleTyrValMetSerLeuGlyAlaIlePheAl 422
|||||
1269 TACGTCTGCGCGACACTTCCACTACGTGCTGCTGCTGCTGCTGCTG 1318
422 alIlePheAlaGlyIleTyrPheTyrMetProLysPheSerGlyAlaVal 439
|||||
1319 GATCTTCCGCGGCTGCTACTGCTGCTCCGAAAGATGACGCGCTCATGT 1368
439 haeProGlyTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGlyAla 455
|||||
1369 ACAAGAGACGCTGGGAGAGCGGCACTTGTGGTACCTTCACTTCCGCG 1418
456 AsnValIleThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProAr 472
|||||
1419 AACCTGGGTCTTCCCGACGACTTCTCCGCGCTGCTGCGCATCCGCG 1468
472 garGlyTrpIleAspTyrProGlyAlaPheAlaLeuTrpAsnLysValSer 489
|||||
1469 CGCGTACGCTGCACTATCCGAGCGCTTGGCGGCGGCAACTGCTGTCTGT 1518
489 eTyrGlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleValIle 505
|||||
1519 CGGTGCGCTCTCATATTCCGCGC.....TTGCGGCTCTCTGATC 1556
506 PheValTyrThrLeuValAlaGlyArgArgGluThrArgPro..... 519

```

```

seq_name: gb_ba:BJA242592
seq_documentation_block:
LOCUS      BJA242592          DNA             30-sep-1999
DEFINITION  Bradyrhizobium japonicum coxB, coxA, coxK, coxL, coxG, coxH,
ACCESSION  AJ242592
VERSION    AJ242592.1
KEYWORDS   assembly protein; coxA gene; coxB gene; coxK gene; coxL gene; coxG
gene; coxH gene; cytochrome c oxidase subunit I; cytochrome c
oxidase subunit II; cytochrome c oxidase subunit III; heme A
synthase; Heme O synthase; hypothetical protein; ORF133; protome
IX farnesyltransferase; shb1 gene; SUR1 homolog; tldd gene; tldd
homolog.
SOURCE     Bradyrhizobium japonicum.
ORGANISM   Bradyrhizobium japonicum.
REFERENCE  1 (bases 1 to 8121)
AUTHORS    Rossmann, R., Lofler, H., Rossi, P. and Hennecke, H.
TITLES     Factors involved in biogenesis of active cytochrome a33 encoded by
the coxB gene cluster from Bradyrhizobium japonicum
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 8121)
AUTHORS    Rossmann, R.
TITLES     Direct Submision
JOURNAL    Submitted (20-MAY-1999) Rossmann R., Eidgenossische Technische
Hochschule Zuerich, Mikrobiologisches Institut, Schmelzbergstrasse
7, CH-8092 Zuerich, SWITZERLAND
FEATURES   location/Qualifiers
            1..8121
            /organism="Bradyrhizobium japonicum"
            /strain="110spc4"
            /db_xref="taxon:375"
            complement(1..440)
            /gene="tldd"
            complement(<1..440)
            /gene="tldd"
            /function="putative role in regulation of carbon storage"
            /codon_start=1
            /transl_table=11
            /product="tldd homolog"
            /protein_id="CAB56816.1"
            /db_xref="GI:6006410"
            /translation="MTNPDTISLDRANILDRDQVREHVARGLAGADGELFLVXSQTE
            ALMFENGRKQATYDTDSQFGLRAVKDQVAVGASDVSLPALIRADAVAAVAGVYS
            GSESAAPPHTNVRILYGDNDPLDAPGFETKVKLAEIDAYLRDXP"
            complement(542..1075)
            /note="ORF177"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="CAB56817.1"
            /db_xref="GI:6006411"
            /translation="WGLSRMARPDSSINALIGLAAALLALPSLAHAGAVSYNG
            DMQIRCDPPGQAQAEQCALIOSYVAEDSNAGLIVYKTDQSRIMRVVAPLGVLL
            PSGLDKINDQVGRAGFVRLPVGCAEVVMDKLGLQLTAKTATFIIIFEPBEGL
            GPPLSLNGLGGYDKLP"
            1523..2362
            /gene="coxB"
            1523..1612
            /gene="coxB"
            /gene="coxB"
sig_peptide

```



```

2593 TCGCCGGGATATCGGTCGCCGATGTCGATCGGATCGCGGACCTG 2642
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 MetasprProgluValGlnTyrMetCysLeuGluAlaArgLeuIleAl 73
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2643 ATGATATCGGCGGTCGACATCTTC..... 2666
73 AspAlaSerGlnThrCysThrAlaAsnGlyHisLeuThrAsnValMet 90
      ..... CACGAGACGACACCTACACAGCTTCG 2694
2667 ..... 2694
90 ALTHrHisGlyIleLeuMetMetPhePheValGlyIleProAlaLeu 106
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2695 TGACCTCCACAGGCGCTGATCATGATCTTCTCATGATCATCGGCGATG 2744
107 PheGlyGlyPheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAs 123
      :|||: ||||| ||||| ||||| ||||| ||||| |||||
2745 ATCGGCGGCTTCGCGACCTGCTGCGCGCTGATGATCGGCGCGCGA 2794
123 pMetAlaPheProArgMetAsnLeuSerPheThrLeuPheIleAla 140
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2795 CATGGCGTCCGCGCATGAAACAATCTGCTGCTGCTGCTGCGCGCT 2844
140 LyrThrAlaMetGlyValAlaSerLeuPheAlaProGlyAspGlyGln 156
      :|||: ||||| ||||| ||||| ||||| ||||| |||||
2845 CCTTCGCGCTGCTGCTGATGTCGACCTTCGTCGAGGCGACCGCGCGC 2894
157 LeuGlySerGlyValGlyTyrValLeuTyrProProLeuSerThrArg. 172
      ||| ||||| ||||| ||||| ||||| ||||| |||||
2895 AACGCGCGCGCGCGCTGCGCATGTCGCGCTGCGCTGCGCTGCGG 2944
173 GlnValGlyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSer 189
      :|||: ||||| ||||| ||||| ||||| ||||| |||||
2945 CCATCCGCGCGCGCGCATGTCGATCGCATCTGCTGCTGCTGCGCGG 2994
189 LysAlaSerSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsn 205
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2995 GTGCTGCTGATCTCGCGCGCATCACTTCATCACACGATCTTCAC 3044
206 MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrp 222
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3045 ATGCGCGCGCGCGCATGACCTGACAGATGCGCGCTTCGTGCTG 3094
222 rIlePheIleThrAlaTyrLeuIleLeuLeuAlaLeuProValLeuAla 239
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3095 GATCTGCTGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3144
239 LysAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePhe 255
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3145 GTGCGATGACATGCTGCTGACCGACCGCAATTTCGCGACGACCTTC 3194
256 AsnProAlaGlyGlyGlyAspProIleLeuTyrGlnHisIleLeuTrp 272
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3195 GCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 3244
272 ePheGlyHisProGlyValTyrIleIleIleLeuProGlyPheGlyIle 289
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3245 CTTCGCGCATCCGCGATGTCATCTGATCTGCTGCGCGCTTCGCGCAT 3294
289 IeSerHisValIleSerThrPheSerLysLysProValPheGlyTyrLeu 305
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3295 TCAGCGAATGCTGTCACCTTCGCGCAAGCGCGCTGCTGCTGCTG 3344
306 PrometValTyrAlaMetValAlaIleGlyValLeuGlyPheValIle 322
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3345 GGCATGCGCTTCGCGCATGTCGCGATCGCGCGCATCGCTTCGTG 3394
332 PAlaHisIleMetTyrThrValGlyMetSerLeuThrGlnGlnSerTyr 339
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3395 GCGCGACACATGTCACGCTGCGCATCTGTCGCGCGACGAGCGCAT 3444
339 hemetLeuAlaThrMetValIleAlaValProThrGlyTyrLysIlePhe 355
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

3445 TCGTCGCGCGACATGATGATCGCGCTTCGACGCGCGCTGAAGATCTTC 3494
356 SerThrPleAlaThrMetTrpGlyGlySerValGlnPheLysSerPro 372
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3495 TCGTGATCGCGCGATGTCGCGCGCTGATGATGATTCGCGCGCGCAT 3544
372 tLeuThrAlaPheGlyPheMetPheLeuPheThrValGlyValAlaThr 389
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3545 GATTCGGCGCGGCTGCTATCTCTCTTCACGTCGCGCTGCGCTACCG 3594
389 LysLeuValLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThr 405
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3595 GCTCGCTGCGCGAAGCGCGCTGACCGCGCTGCTGCTGCTGCTGCTG 3644
406 TyrValValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePhe 422
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3645 TACGTCGCGCGCATTCACATGTCGTCGCTGCTGCTGCTGCTGCTG 3694
422 AliePheAlaGlyIleTyrPheTyrMetProLysPheSerGlyArgAla 439
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3695 GATCTTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3744
439 heProGluThrAlaAlaLysLeuHisPheThrThrPhePheIleGlyAla 455
      :|||: ||||| ||||| ||||| ||||| ||||| |||||
3745 ACAACGAGACGCTGCGAAGCGCGCTGCTGCTGCTGCTGCTGCTG 3794
456 AsnValThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetPro 472
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3795 AACCTGCTGCTTCTCCGCGACCTTCCTGCTGCTGCTGCTGCTGCTG 3844
472 GATGlyTyrIleAspTyrProGlnAlaPheAlaLeuTyrAsnLysValSer 489
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3845 CCCTGATGCTGATTCGCGACGCTGCTGCTGCTGCTGCTGCTGCTG 3894
489 eTyrGlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleValIle 505
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3895 CGCTGCTGCTTCATTCATTCGCGC.....TTGCGCGCTGCTGCT 3932
506 PheValTyrThrLeuValAlaGlyArgGlnTyrArgPro..... 519
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3933 TTCCTCATTCGCTGATTCGATGCTTCGCAAAAGCTGCGCTGCGCA 3982
520 AsnProTrpGlyGlnPheAlaAspThrLeuGlnTyrThrPheLeuSer 536
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3983 CAATCGTGGGCTGCGCGCGCGACGCTGAGTGAAGCTGCTGCTGCTG 4032
536 roProProAlaHisThrPheGlyThrLeuProLys 547
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
4033 CGCGCGCTTCATTCATTCGAGTTCGAGTCTGCTGCTGCTGCTG 4067
seq_name: gb_ba:BJU33883
seq_documentation_block:
LOCUS BJU33883 31495 bp DNA BCT 10-AUG-2000
DEFINITION Bradyrhizobium japonicum putative epoxide hydrolase Ephb (ephb),
putative stress-induced protein Ohr (ohr), putative transmembrane
transcriptional regulator protein TtrR (ttrR), putative
two-component system transcriptional response regulator Tcsr
(tcsr), integral inner membrane metabolite transport protein MtdA
(mtdA), orf157, and orf235 genes, complete cds; Fo ATPase operon,
complete sequence; probable acetyltransferase Pact (pact),
mitochondrial processing peptidase-like protein Mpp (mpp),
threonine synthase Thrc (thrc), two-component system
transcriptional regulator ElmR (elmR), putative kinase MtdB (mtdB),
and orf13 genes, complete cds; cox operon, complete
sequence; orf177, TldD (TldD), signal peptidase Sips (sips), and
orf175 genes, complete cds; and putative glutathione synthetase
GshA (gshA) gene, partial cds.
ACCESSION U33883
VERSION U33883.2 GI:8708897
KEYWORDS
SOURCE Bradyrhizobium japonicum.

```

ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Bradyrhizobium group; Bradyrhizobium.
REFERENCE 1 (bases 1 to 31495)
AUTHORS Muller, P., Ahrens, K., Keller, T. and Klauke, A.
TITLE A TnpA insertion within the Bradyrhizobium japonicum sips gene, homologous to prokaryotic signal peptidases, results in extensive changes in the expression of PBM-specific nodulins of infected soybean (Glycine max) cells
JOURNAL Mol. Microbiol. 18 (5), 831-840 (1995)
MEDLINE 96422470
PUBMED 8825087
REFERENCE 2 (bases 1 to 31495)
AUTHORS Mueller, P.
TITLE Extended sequencing of a DNA fragment of B. japonicum adjacent to the cox operon
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 31495)
AUTHORS Mueller, P.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-1995) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany
REFERENCE 4 (bases 1 to 31495)
AUTHORS Mueller, P.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Frisch-Str., Marburg 35032, Germany
REMARK COMMENT Sequence update by submitter
FEATURES On Jun 26, 2000 this sequence version replaced gi:1200339.
source location/Qualifiers
1..31495
/organism="Bradyrhizobium japonicum"
/strain="USDA1105pc4"
/db_xref="taxon:375"
/clone="C27"
complement(7..909)
/gene="ephB"
complement(7..909)
/gene="ephB"
/codon_start=1
/transl_table=11
/product="putative epoxide hydrolase ephB"
/protein_id="AAF78792.1"
/db_xref="GI:8708901"
/translation="MKPFTIKQIDAGVSMAMRGVPSNGVPAICCTAGPYDIHAFV
VAPDGKGRVYIIPYLRGTTHFLSSTPRQARAPAMAVDIIALMDKIDIKAV
AGFDGARTADIIAALPDAARMCRSAGYLISQAGSAPLPSPAEIQMWQYFAT
EGRAGYEKTYHDFAKLWKLASPMKEDATFDSAAALDNKDHVAITIHNYRRLG
LAKGEAKYELEKTKLAAPLISVPTITMEGANGAPHPDPSTYAKKFGRHXYRLITG
GVGNLPRAPQAPAKAVTDACGA"
complement(1071..1490)
/gene="Ohr"
complement(1071..1490)
/gene="Ohr"
complement(1071..1490)
/gene="Ohr"
/transl_table=11
/codon_start=1
/product="putative stress-induced protein Ohr"
/protein_id="AAF78793.1"
/db_xref="GI:8708902"
/translation="MTQAAKLLYTARTHTGGGRODOMSRSSDGRILDVRLSPGAGIG
TNEOLFAGWSACFEGAMEIARRKRTILPKKCAIDAEIDLLDEGAYSLRRLNVS
LGLDREVARDDIIDEHQTCTPSKAVRGNDIVAVALT"
1690..4398
/gene="ttr"
complement(1690..4398)
/gene="ttr"
complement(1690..4398)
/transl_table=11
/codon_start=1
/product="putative transmembrane transcriptional regulator protein Ttr"
/protein_id="AAF78794.1"
/db_xref="GI:8708903"
/translation="MTPEAPAPVAGPLSGFPTVIERLVTRDGVALLPGTKAPFLI
ALMSRNEVYSSKMDIALVMPGMAVEANIREVVALRKLVDGKGAARYITLLSGG
YCFVAPISQTDVPAERPPAGMCPVVKPNRLQRVNGDEVYAAISDKLITSHDVTI
VGGGQKTAIVAAVAHDDLETFADAAHVDLAUSDPLVLTITLLMGLRSHDPT
VAGGCRISQTRSCCLLNDCEHVAAPLAEIFHAPPHHLLATSERLRVEQ
VYRLADLAPDDAGLAAVQYPAIQFLERATVGAQOAILDDNAAIVGILRSV
DNMALEIAGRAVEAYGLEOTALLDERLILMOGOTRAPROKTIQATLMSGSL
SCHEVLRRLVAFAGHTIDALEVYPPDRVDRARLPDAISLVAKSWAPRRGAM
MYRLDITTRAVILEIPEPDALAGRORLPAMAGTGRHDMCNADADEAOSTEAL
HNVRALDWCFGANGDCTGIALAAAPLFLTMSLLTECRRESEALLAIDPSTRG
AEEMHIOALGILMTETGRQSGARRIDQPCDRGARRANQDLQRMHIFERM
GOFDAALGYAROSLTVAETLGDAAASLGLASLGLVHLAGEHRLNMLEAAWQPG
TERSTVHGDRNRAGISLARLMLQGRGDAQLAROTVEAQMHPITLCTALI
WLVSIDMSGLDDESIDHPITAHATSRMGPLYLSGVGKGEALIRGDAAGVET
IRACIRLHDAGYEILTPFNALVHGILALGOVQASRLIDARLRLVGGDHLVMP
ELLRMRKVLISLPEPRTGAEFTCMQSLSLSGQCESLGAAGP"
2158..2217
/gene="ttr"
/note="putative"
/note="transmembrane-region site"
complement(4540..5211)
/gene="tcsr"
complement(4540..5211)
/gene="tcsr"
/codon_start=1
/transl_table=11
/product="putative two-component system transcriptional response regulator Tcsr"
/protein_id="AAF78795.1"
/db_xref="GI:8708904"
/translation="MNAFSTHLYIADHPLEFRLDAQNAVASVLTARIDAGSFEDLT
KLROTSDVDLIDLSMPGISGSLIYRAQYPAIPVIVASDSATIRSLDFG
ASGFIKPRFGEVSRDALIKVMGDWVPADTQALADPDMDRLDRIVTLTPOOVR
VLMHLSGLNKQIATYELGSEATIKAHVSAIIQKLGVESTRQAVLAARLAGGQKQ
GTSFG"
complement(5050..5052)
/gene="tcsr"
/note="putative; phosphorylation site"
5518..7185
/gene="mtbA"
5518..7185
/gene="mtbA"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="integral inner membrane metabolite transport protein MtbA"
/protein_id="AAF78796.1"
/db_xref="GI:8708905"
/translation="MSTMAVSOAGVGMTRDERFVLIASSLGVFEWYDFLYGLSAS
IIGAFPSAPYPRATDIFALIAFAAGFLVPEFGAIVGRIQDIVGRKFTLVILING
LSTFIVGLIPAAATIGIAPIITLIALRLRQGLAGGYGAATVYVAHPNKGKRY
TSPTQTRATGICPLSLIVITPFRSALGEADPAAMGWIPLVSVLLIGISVRLRLN
ESPTFOKKDGGKSSKAPLTFARFNMQNKILVALLGLGVGQGVVYVQFVAPLFL
QSILIKVGYTANLLIAMSLLGTGFTVFGMLSKITRKP11IGGCLIALTTPPLTK
MITSNANPALEKAEIEQGRGGRSGRLRRVQGRYRLRLARLHRTPLRPVVGQV
FDIDGGRFREDGROGSGCLREBQSCGSRGRRLRSEDRRRRHHEDASVH
LPSAAVATIGLFTLVIFVWYGPILAMIVELPRTIRITMSLPYHINQWFGGL
PATAPAVASGDIYAGIMYVIVASTIVVIGFLFLETRDVDIKA"
complement(7350..7823)
/gene="Orf157"
complement(7350..7823)
/gene="Orf157"
/note="Orf157"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="unknown"
/protein_id="AAF78797.1"
/db_xref="GI:8708906"
/translation="MTSARSVTPIGAILIGASHHYLIVAHRGDGARGLHGPDPPTAA
IAEARARGAVTLPIRHTSASLTIOENNDPVLVDITLALSLAPENAGMTHPDECPD

gene
DMPGH1KTLTGTSTSLQVPELNGRLALGTWQATILYLEHRAHPHREVVLLQFIGANG"
complement(7853. .8560)
/gene="orf235"
complement(7853. .8560)
/gene="orf235"
/note="ORF235"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="unknown"
/protein_id="AAE78798.1"
/db_xref="GI:8708907"
/translation="MPCGIADVPVPARRGGECDTSDVYVHNIMAFGSGAGGDDCOORR
SNCRIATARRRRKRPLOHDADPPRAGASTPAPAVDILQERRGILLGARACRRR
RRPRAPDSREAREGTCARGCTDRYKIDALPKSDPLKADKSPATODRPPAERPR
RRSCVGVPEKGLADHRGQPPRAGSVEDELSSWRKPGHIRKGLSVHSASALPPPA
VTRIGNOSWLECAIAH"
8893. .9240
/gene="uncI"
/note="orf115"
8893. .11493
/gene="Fo ATPase operon"
/note="constitutively expressed"
8893. .9240
/gene="uncI"
/note="ORF115; initial gene of the Fo ATPase operon"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="unknown"
/protein_id="AAE78799.1"
/db_xref="GI:8708908"
/translation="MEINIDRRPVRRFPNGSEILLSCPNFAATSPKSPACDGDGR
AARAAAMALGRLSSELVAGVVGAGIGWGFDRLLSTSPGFIIVLLLGIVAGVNVV

```

Quality: 1829.00      Length: 545
Ratio: 4.073          Gaps: 4
Percent Similarity: 82.385 Percent Identity: 61.83555
alignment block:
US-09-712-766-2 x BJ033883/rev ..

```

Align seg 1/1 to reverse of: BJU33883 from: 1 to: 31495

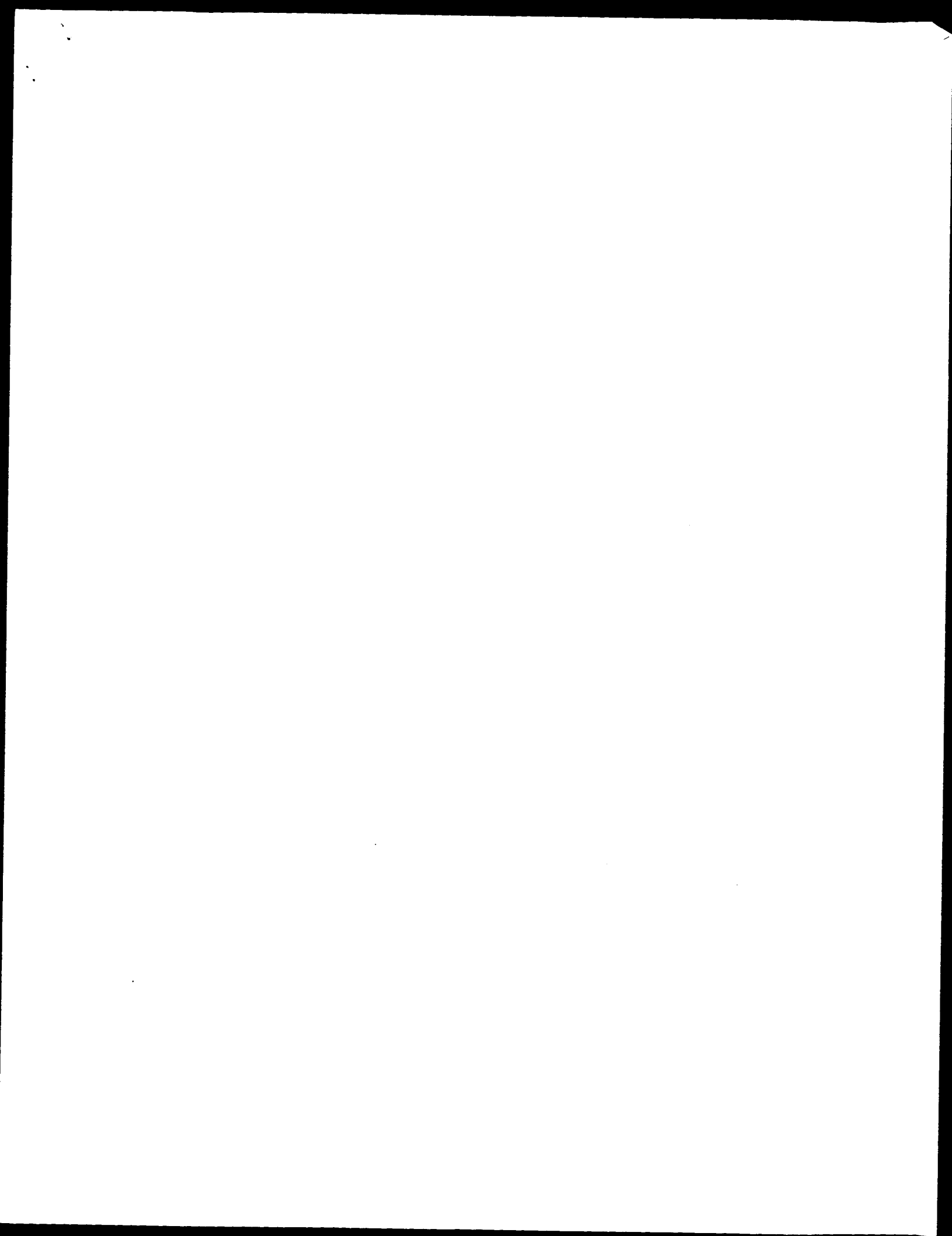
7	hlsqlyhnsiaspnhisnsluysgnglypnehrthArtrpHEME	23
24557	CAGGACATGACGAGCAGCCCATCCGACCGGATGGCCAGCTACTCTTA	24508
23	cserthrAsnHsLysAspIleGlyLeuLeuTyrlLeuValAlaIaGlyV	40
24507	TTGCAACCAACCAAGAGCATGGCGACGATGACCTGATCTTGGGGTCA	24458
40	aIValGIyrlhelleserValIleuphethValTyrlmetrgLeuIleu	56
24457	TGCGCGGCTCATCGGTGGCGGATGTCATGCGCATGCGTCCGACGCG	24408
57	metAspProGlyValAlGlnTyrmetCysLeuGlnGlyAlaIargLeuIleAl	73
24407	ATGTATCCGGCGGTCAAGATCTTC.....	24384
73	aAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMetV	90
24383CAGGACGCGACACACTACACAGCTTTCG	24356
90	aIThrTyrlHisGlyIleLeuMetMetPhePheValGlyIleProAlaLeu	106
24355	TCACCTCCACAGCGCTGTATCATGATCTTCTCATGTATGATGCGGCGATG	24306
107	PheGlyIlyPheGlyAsnTyrlLeuMetProLeuGlnIleGlyAlaProAs	123
24305	ATGCGCGGCTTCGCACTGGTCTTCGCGCGTGAATGATGGCGGCCCGCA	24256

123	pmctAlaPheProArgMetAsnAsnLeuSerPheTrpLeuPheIleAlaG	140
24255	CATGGCGCTTCCCGCGCATGACACACTCTCTTGGCTGGCGGCGCT	24206
140	LYTHAlaMetGlyValAlaSerLeuPheAlaProGlyLysArgIleGln	156
24205	CCCTTCGGGCTGCTGCTGATGTGTCACCTCTGTCGAGGCGAGCCGGGCGCC	24156
157	LeuGlySerGlyValGlyIleProGlySerPheProPheSerThrArg..	172
24155	AACGGCTGGGGGCGCGCGTGGACCAATGACGTCCGCTGTCCAGCTCCGG	24106
173	.GluAlaGlyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerG	189
24105	CCATCCCGGCGCGCGAGCTGGACCTGGCATCTCTGTCCGTCGATCTGGCGG	24056
189	LYAlaSerSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsn	205
24055	GTCCCTCGTGCATCTCCCGCGCATCACTCATCACCACGATCTTTCAAC	24006
206	MetArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSe	222
23955	GATCCGCGTGGACCGTCTCTCTGCTGTGTTCGCTGGCGGTCGCGCG	23906
239	LYAlaIleThrMetLeuLeuThrAspArgAspPheGlyThrThrPhePhe	255
23905	GTCCGATCAACATGCTGCTACCGACCGCAATTTGGACGACCTTCTTC	23856
256	AsnProAlaGlyIleGlyLysAspProIleLeuTyrGlnHisIleLeuTrpPh	272
23855	GGCGCCGACGGCGCGGCGACCCCGTCCGTCGACGACCTGTCTGGTT	23806
272	ephGlyHisProGlyValTyrIleIleLeuProGlyPheGlyIleI	289
23805	CTTCGGCATCCCGAGGTGATACCTGATCTCGCCGCTTCCGGCAAGA	23756
289	IeSerHisValSerThrPheSerLysLysProValPheGlyTyrLeu	305
23755	TCAGCGACATGCTGCCAATCTTCGGCGCAACCCGCTTGGCGCATCTC	23706
306	ProMetValTyrAlaMetValAlaIleGlyValLeuGlyPheValValTr	322
23705	GGCATTGGCTACGCAATGATGTCGCGCATCCGGCGCATCGCTTCGTGGTGTG	23656
322	PAHisHisMetTyrTrpValGlyMetSerLeuThrGlnGlnSerTyrP	339
23655	GGCGACACACATGTACAGGCTGGCATGTCTCGCGGACGACGAGGCTTAAT	23606
339	hemeLeuAlaThrMetValIleAlaValProThrGlyIleLysIlePhe	355
23605	TCGTGCGCGCGACGATGGTGAATCGCGGTCCGACCGCGGTGAAGATCTTC	23556
356	SerTrpIleAlaThrMetTrpGlyLysValGlnIleLysSerProMe	372
23555	TCGTGATATCCGACATGTGGGGCGCGCTGATTCGAAATTCGCCGCGCAT	23506
372	IleuTrpAlaPheGlyPheMetPheLeuPheThrValGlyLysValThrG	389
23505	GATTCGGGCGCGGCTTCATCTTCCTTCACGGTGGGTGGCGTCAACG	23456
389	LYIleValLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyr	405
23455	GCCTGTCTCTGGCAACCCGCGCTCGACCCCGTCTCCAGAGAACCTTAC	23406
406	TyrValValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAl	422
23405	TACGTCGTCGGCGCACTTCCACTAGCTGTGTGCGCTCGGTGCAAGTGTTCGC	23356
422	AllePheAlaGlyIleTyrPheTyrMetProLysPheSerGlyArgAlaP	439


```

431 tProLysPheSerGlyAlaGlyAlaPheProGluTrpAlaAlaLysLeuHis 448
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7601 TGGAAAAAATCTGTGGTCTTCATATGTCTGAATTTTGGGACAAATTCATT 7650
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
448 heTrpThrPhePheIleGlyAlaAsnValThrPhePheProGluHisPhe 464
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7651 TTTGGCTTTTCTTTGTAGGGGTATATTACATTTTTCCTATGCATTTT 7700
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
465 LeuGlyAlaGlnGlyMetProArgArgTyrIleAspTyrProGluAlaPhe 481
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7701 TTAGGATTTTCGGAAAGCCAGACGAATTCGCCATTTATCTGTAGTCTT 7750
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
481 eAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaAs 498
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7751 TGTGTGATGTGAATGCAGCTGTGATGTAGGTACGAGATCTATCTTCAGTCT 7800
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
498 ePheLeuPhePheIleValIlePheValTyrThrLeuValAlaGlyArg 514
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7801 GACCTTATATTTTTTTCATGTGA..GTTTACGATACATTAAGTGGAGAA 7847
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
515 ArgGluThrArgProAsnProTrp.....GlyGluPheAlaAs 527
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7848 GAAACATGTCCTAAATAATCTCTTGGGAAACAGTCCAGGAACCTCGCGG.. 7895
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
527 pThrLeuGluTrpThrLeuProSerProProAlaHisThrPheGlu 544
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
7896 .ACATTAGAGTGGTTAACTCGTCACCAACCAAGCTTACCATCTTTTGAAG 7944
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
544 hrLeuPro 546
      ::::: :::::
7945 AAGTTTCCA 7952

```



Align seg 1/1 to: AAA91489 from: 1 to: 1674

```

1 MetAlaAspAlaAlaIleHisGlyHisAspHisHisGlnGlnGlnGlyPhe 17
1 ATGGCAGACGCCGCCATTCACAGCCATGACCACCATGAGCAAGCAAGCGTT 50
17 ePheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeuLeu 34
51 CTTCACCGCGGTGGTTTCATGTCGACCAACACCAAGAAGCATCGCTCTCAT 100
34 YrLeuValAlaAlaGlyValAlaGlyPheIleSerValLeuPheThrVal 50
101 ACCTTGTAGCGGCTGCTGTTGTTGTTTCATTTCCGCTCCGTTCCACCGTC 150
51 TyrMetArgLeuGluLeuMetAspProGlyValAlaGlnTyrMetCysLeu 67
151 TACATGGCGCTTGAAGTATGATCGGGGATCCGATCATATGATGCTTGA 200
67 uGlyAlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnGlyH 84
201 AGCGCGACGCTGTATCGCGGATCCCTCGCAGACATGTACGGCAACGGAC 250
84 IseuThrAsnValMetValThrTyrHisGlyIleLeuMetPhePhe 100
251 ACCGTGGAACGTCATGTTACCTACCATGATTCGTATGATGTTCTTT 300
101 ValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetProLe 117
301 GTGGGTATCCCGCATTTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 350
117 uGlnIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeuSer 134
351 GCAAATCGCGCGCTCCGATATGCTCCCGGTATGACAAACCTGTCGT 400
134 hePheLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAla 150
401 TCTGGTGTTCATTCGCGTACCGGATGGCGGTGGCTTCGCTGTCGCA 450
151 ProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTyrPheValLeuTyrPr 167
451 CGGGCGGTCGACGCTGACGCTGGGTTGGGGCTGGGTTGGTGGTGGTGG 500
167 oProLeuSerThrArgGluAlaGlyTyrSerMetAspLeuAlaIlePheAla 184
501 GCCGCTGTGACCCCGCAAGCTGGGTATTCATGACCTCCGACCTTTTCG 550
184 IValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleAsnMetIle 200
551 CGGTTCACTGTGCGGTGCTTCCTGATCATGCGCGCATCAACATGATC 600
201 ThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisLysValPr 217
601 ACGACCTTCCTGAAACATGCGCGGCCCGCGCATGACCTGCACAAAGTGCC 650
217 oLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLeuLeuAlaAla 234
651 GTTGTCTCGGTGTCGATCTTATCAGCGCTGATGATGATGATGATGATG 700
234 euProValLeuAlaGlyAlaIlePheMetLeuLeuThrAspArgAsnPhe 250
701 TGGCGGTTCTGGCTGTCATCACTGCTGCTGACCGACCGTAACTTTC 750
251 GlyThrThrPhePheAsnProAlaGlyGlyGlyAspProIleLeuTyrG 267
751 GGCACGACCTTCTTCAATCCTGCTGGCGGGGCGGACCCGATTCGTACCA 800
267 nHisIleLeuThrPhePheGlyHisProGlyValTyrIleIleIleLeuP 284
801 ACACATCTGTGTGTTCTTGGGACCCGGAAGGTACATCATCATCATCTGC 850
284 roGlyPheGlyIleIleSerHisValValSerThrPheSerLysLysPro 300

```

```

851 CGGCTTTGGCATCATCAGCCATGCTGTCGACCTTCTCGAAAAAGCCG 900
301 ValPheGlyTyrLeuProMetValTyrAlaMetValAlaIleGlyValLe 317
901 GTCTTCGTTACCTCGCGATGCTCTATGACATGTCGCAATGCGGTTCCT 950
317 uGlyPheValValTrpAlaHisMetTyrThrValGlyMetSerLeu 334
951 GGGCTTTGTGCTGGCGCACCATGTAACCGTTGATGTCGCTGA 1000
334 hArgInsInsTyrPheMetLeuAlaThrMetValIleAlaValProThr 350
1001 CCCAGCAATCTTACTTCATGCTGCGCACCATGTAATCGCGTCCGACAC 1050
351 GlyIleLysIlePheSerTrpIleAlaThrMetTrpGlyGlySerValG 367
1051 GGGATTAAGATCTTCTGTCGATGCGCCAGCATGTGGGGGCTCGCTTGA 1100
367 uPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPheThrV 384
1101 GTTCAATCGCGCATGCTCTGCGCTTGGCTTATGTTCCGTTCCACCG 1150
384 AlGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAspArgAla 400
1151 TGGGTGTGTGACCGGTATGTCGTGGCCCAAGCGGCTGTGACCGTGA 1200
401 TyrHisAspThrTyrTyrValValAlaHisPheHisTyrValMetSerLe 417
1201 TATCAGCACCATCTTATGCTGTCGCGCACCTTCATATGATGATGCTGT 1250
417 uGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetProLysP 434
1251 GGTGTGATCTTGGCATCTTCGCGGTATGCTGCTGCTGCTGCTGCTGCT 1300
434 heSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTrpThr 450
1301 TCTCGGCGCGCTTCCCGAATGGCGTCCAAAGCTGCACTTGTGACCC 1350
451 PhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuGlyAr 467
1351 TTCTTCATCGGTGCGCAACGTCACGTTCTTCCCGACGACCTTCCGAGAC 1400
467 gGlnGlyMetProArgArgTyrIleAspTyrProGluAlaPheAlaLeu 484
1401 TCAGGTATGCGCGCGCTTACATCGACTATCCCGAAGCTTGGCGCTGT 1450
484 rPAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPheLeu 500
1451 GGACAAAGTCTGCTGCTATGTCGCTTCCGCTTCCGCTTCCGCTTCCG 1500
501 PhePheIleValIlePheValTyrThrLeuValAlaGlyArgArgGlyH 517
1501 TTCTTCATCGTATCTTGTCTATACGCTGTTGCTGCGCGCGCGACGAC 1550
517 rArgProAsnProTrpGlyGluPheAlaAspThrLeuGluTrpThrLeuP 534
1551 CCGTCCACACCGGTGGGCGCAATTCGCGCATACGTCGAATGAGACGCTGC 1600
534 roSerProProProAlaHisThrPheGluThrPheLeuProLysArgSerAsp 550
1601 CATCAGCGGCTCGCGCGCCACACGTTGCGAAAGCTGCGCAAGGCTCGGAC 1650
551 TrpAspLysHisProSerHis 557
1651 TGGGACAAGCATCCCTCGCAC 1671

```

seq_name: /sids8/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV53011

seq_documentation_block:

ID AAV53011 standard: DNA: 1693 BP.

XX AAV53011:
XX
XX


```

859 GAAAAAAGAACCATTTGGATACATAGCTGCTGAGCTATGATATCA 908
314 IleglyValLeuGlyPheValValTrrPalaHisMetLysThrValGI 330
909 ATTGGATTCCTAGAGTTTATCTGTGACACACCATATATTTACAGTAG 958
330 yMetSerLeuThrGlnInserTyrPheMetLeuAlaThrMetValILa 347
959 AATGAGAGTACACACAGCATATTTTCACCTCCGGTACCATATATATG 1008
347 IAlaProThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGly 363
1009 CTATCCACCACCGGCTCAAAAGTATTTAGCTGATGCGCACACCTCC 1058
364 GlySerValGluPheLysSerPrometLeuTrrPalaPheGlyPheMet 380
1059 AGCAATATGAAATGATCTGCTGACATGCTGAGCCCTAGATCATCTT 1108
380 eLeuPheThrValGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyL 397
1109 TCTTTTCACCGTAGTGCTGCTGACATGCTGATTTAGCAAACTCATC 1158
397 euAspArgAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyr 413
1159 TAGACATCGTACTACACGACAGCTACTAGCTTACGCCACTTCCACTAT 1208
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyr 430
1209 GTCTCTCATATGAGAGCTGATTTGCCATCATAGAGAGCTTCATTCAC 1258
430 rMetProLysPheSerGlyArgAlaPheProGluTrrPalaIleLysLeu 447
1259 ATTTCCTCATTTCTCAGGCTACACCTAGACCAAACTACGCCAAATACC 1308
447 IAspThrTrpPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
1309 ATTTCATCTCATTTCTATTCGCGGTAATCTCACTTCTTCCACACAC 1358
464 PheLeuGlyArgGlnGlyMetProArgTyrTyrIleAspTyrProGluAl 480
1359 TTTCTCGGCTATCCGGAATGCCCGAGCTTACTCGAGCTACCCGATGC 1408
480 aPheAlaLeuTrrPalaHisValSerTyrGlyAlaPheLeuAlaPheA 497
1409 ATACACCATGAAACATCTATCATCTGTAGGCTCATTTCTCTCTAA 1458
497 IAspPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
1459 CAGCA...GTAATATTAATATTTTCATGATTTGAGAACCTTCGCTCG 1505
514 ArgArgGluThr.....ArgProAsnProTrrPalaGluPheAl 526
1506 AAGCGAAGAGCTTAATATGAGAACACCC.....TC 1537
526 aAspTrpLeuGluTrrPheTrpLeuProSerProProAlaHisThrPheG 543
1538 CATTAACCTGAGTGAATATGATGATGCCCCACCTTACCAACATTCG 1587
543 Lu 543
1588 AA 1589

```

seq_name: /SID8/gcgdata/geneseq/NA1995.DAT:AA705905

seq_documentation_block:

ID AA705905 standard; DNA: 1738 BP.

AC AA705905;

29-MAY-1996 (first entry)

DE Mitochondrial cytochrome c oxidase subunit I DNA.

```

KW Human; mitochondrial cytochrome C oxidase; COX; subunit I; subunit II;
KW subunit I1; mutation; Alzheimer's disease; AD; sporadic form;
KW diabetes mellitus; IDW; ss.
OS Homo sapiens.
EH Key
EH CDS
FT Location/Qualifiers
FT 72..1613
FT /tag= a
FT /product= COX C subunit I
FT 1..71
FT /tag= b
FT 1614..1738
FT /tag= c
FT 147..149
FT /tag= d
FT /number= codon 26
FT /note= "Position of GCT>ACT, Ala>Thr mutation in AD patients"
FT 225..227
FT /tag= e
FT /number= codon 52
FT /transl_except= pos: 225..227, aa:Tyr (possible mutant)
FT /note= "Position of CAC>CCC, His>Tyr; CAC>CTC, His>Leu mutation in AD patients"
FT 267..269
FT /tag= f
FT /number= codon 66
FT /note= "Position of ATC>GTC, Ile>Val; ATC>ACC, Ile>Thr mutation in AD patients"
FT 321..323
FT /tag= g
FT /number= codon 84
FT /note= "Position of CCC>CTC, Pro>Leu mutation in AD patients"
FT 333..335
FT /tag= h
FT /number= codon 88
FT /note= "Position of GGT>GAT, Gly>Asp mutation in AD patients"
FT 378..380
FT /tag= i
FT /number= codon 103
FT /note= "Position of TGA>CGA, Trp>Arg mutation in AD patients"
FT 396..398
FT /tag= j
FT /number= codon 109
FT /note= "Position of CTC>CCC, Leu>Pro mutation in AD patients"
FT 402..404
FT /tag= k
FT /number= codon 111
FT /note= "Position of CTC>CCC, Leu>Pro mutation in AD patients"
FT 477..479
FT /tag= l
FT /number= codon 136
FT /note= "Position of TAC>CAC, Tyr>His mutation in AD patients"
FT 534..536
FT /tag= m
FT /number= codon 155
FT /note= "Position of GTC>ATC, Val>Ile; GTC>GCC, Val>Ala mutation in AD patients, position of mutation hotspot"
FT 570..572
FT /tag= n
FT /number= codon 167
FT /note= "Position of ACA>GCA, Thr>Ala mutation in AD patients, position of mutation hotspot"
FT 579..581
FT /tag= o

```


[illegible]

1109	TCCTTTACACCTAGGTGGCCCTGACCTGGCATGTGATTACCAAACTCATCAC	1158
397	euAspArgLalaTyrHisAspThrTyrTyrValValAlaHisPheHisTyr	413
1159	TAGACATTCGACTACACGACACGACTACGTTGAGCCCATTCACCTAT	1208
414	ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyr	430
1209	GTCCTACCAATAGAGCGTGATTGGCATCATAGAGGCTTCATTCACTG	1258
430	rMetProLysPheSerGlyArgAlaIlePheProGluTrrPheAlaLysLeuH	447
1259	ATTTCCTCATTTCTCAGGCTACACCTAGACCAAACTACGCCAAATCC	1308
447	IsPheTrpThrPhePheIleGlyAlaAsnValThrPheProGluHis	463
1309	ATTTCATCATCATATTCATGCGGCTAATCTAAGTTCTTCCCAACAC	1358
464	PheLeuGlyArgGlnGlyMetProAlaArgTyrIleAspTyrProGluAl	480
1359	TTTCTCGGCTATCCGGAAATGCCCGACGTTACTCGGAGTACCCGATGC	1408
480	apheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheA	497
1409	ATTCACGACATGAAACATCTCATCATCTTGAGGCTCATTAATTTCTTAA	1458
497	laserPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly	513
1459	CACGCA...GTAAATATTAAATATTTCATGATTGAGAAAGCTTCGGTTCG	1505
514	ArgArgGluThr.....ArgProAsnProTrrPheGluPheAl	526
1506	AAACGAAAAGTCTAATAGTAGAAGAACCC.....TC	1537
526	aaSprThrLeuGluTrrThrLeuProSerProProAlaHisThrPheG	543
1538	CATTAACCTGAGTACGACTATATGATGATGCCGCCACCCCTACACACATTCG	1587
543	lu	543
1588	AA	1589
seq_name: /SID8/gcdata/geneseq/geneseqn/NA1999.DAT:AA211492		
seq_documentation_block:		
ID	AA211492 standard; DNA; 16569 BP.	
XX		
AC	AA211492;	
XX		
DT	02-NOV-1999 (first entry)	
XX		
DE	Mutated human mitochondrial DNA sequence.	
XX		
KW	Mutation: human; mitochondrial gene; abnormality detection;	
RW	diabetes mellitus; ds.	
XX		
OS	Homo sapiens.	
OS	Synthetic.	
XX		
Key	Location/Qualifiers	
FT	3423	
FT	mutation	
FT	/*tag= a	
FT	/note= "native G is mutated to T"	
FT	mutation	
FT	3426	
FT	/*tag= b	
FT	/note= "native A is mutated to G"	
XX		
PN	JP11221077-A.	
PD	17-AUG-1999.	
XX		
DE	04-FEB-1998; 98JP-0023196	


```

5957 ATACCTATTATTCGGCGCATGAGCTGAGTCCCTAGACACCTTAACCC 6006
50 a1TyrMetArgLeuGluLeuMetAspProGlyValGlnTyrMetCysLeu 66
6007 TCCTTATTCGAGCGGAGCTGGCCAGCCAGGC..... 6038
67 GluGlyAlaIArgLeuIleAlaAspIAserGlnThrCysThrAlaAspG1 83
6039 .....AACCTTCAGGTAAACGA 6055
83 yHISLeuTrpAsnValMetValThrTyrHisGlyIleLeuMetMetPheP 100
6056 CCACATCTACACAGCTTATGCTACAGCCCATGATGATTAATATCTTCT 6105
100 heValGlyIleProAlaLeuPheGlyIlePheGlyAsnTyrLeuMetPro 116
6106 TCATGATATACCCATCATATCGAGGCTTTGGCACTGACTAGTTCCTCC 6155
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeuSe 133
6156 CTAATAATCGGTCGCCGATATGGCTTTCCCGCATTAACACATACAG 6205
133 PheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaIAserLeuPheA 150
6206 CTTCGACTCTTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6255
150 IapProGlyIAspGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 166
6256 TG.....GAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6290
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181
6291 CTTCCCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6340
181 a1IlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIle 198
6341 CACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6390
198 smetIleThrPheLeuAsnMetAlaGlyAlaProGlyMetThrLeuHis 214
6391 ATTTCATCTACACATATTAATTAATAAACCCTGCCATTAACCAATAC 6440
215 LysValProLeuPheSerIlePheIlePheIlePheIlePheIlePhe 231
6441 CAACGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6490
231 IleValAlaLeuProValLeuAlaGlyAlaIlePheMetLeuLeuAspA 248
6491 CCTATCTCTCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6540
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyIleGlyIleAsp 264
6541 GCAACCTCAACACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6590
265 LeuTyrGlnHisIleLeuTrpPhePheGlyHisProGlyValTyrIle 281
6591 CTATACCAACACCTATCTGATTTTTCGGTCAACCTGAAGTTTATATCT 6640
281 e1IleLeuProGlyPheGlyIleIleSerHisValIAserThrPheSer 297
6641 TATCTTACACAGGCTCGAATAATCTCCCATATTTTATTAAGTACGCG 6690
298 ..LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
6691 GAAAAAACAACCATTTGGATATACATAGGTATGCTGAGCATATGATCA 6740
314 I1eGlyValLeuGlyPheValValTyrAlaHisMetLeuTyrThrValG1 330
6741 ATTTGCTTCTCTAGGCTTATCTGAGGACACCATATTTATTTACAGTAG 6790
330 yMetSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleA 347
6791 AATAGACTATAGACACAGGATATTTTCACTCCGCTACCATTAATCATCG 6840

```

```

347 IalValProThrGlyIleLeuSerIlePheSerTrpIleAlaThrMetTrpGly 363
6841 CTATCCCAACCGCGGCTCAAAAGTATTTAGTGAAGTCTCCACACCTCCACG 6890
364 GlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetPhe 380
6891 AGCAATATGAAGATGATCTGCTGAGTGTGCTGAGGCTTCAATATCTTCT 6940
380 eLeuPheThrValGlyValThrGlyIleValIleValAlaGlnAlaGly 397
6941 TCTTTTACCGCTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6990
397 euAspArgAlaIleTyrHisAspThrTyrTyrValValAlaHisPheHis 413
6991 TAGACATCGTACTACACGACAGACGACTAGCTAGGCTTCAAGCTTCCACT 7040
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPhe 430
7041 GTCCATCATATGAGAGCTGTATTTTCCCATATAGGAGCTTCAATTCAC 7090
430 rMetProLysPheSerGlyAlaGlyAlaPheProGlyTrpAlaAlaLys 447
7091 ATTTCCTATTTCTCAGGCTACACCTTACGACCAACCTACGCCAATAATCC 7140
447 IAsPheTrpThrPhePheIleGlyValAlaAsnValThrPhePheProGln 463
7141 ATTTCATCATATGATTCATGCGGCTTAATCTAATCTTCTTCCACACAC 7190
464 PheLeuGlyAlaGlnGlyMetProAlaGlyTyrIleAspTyrProGln 480
7191 TTTCTCGGCTATCCGGAATGCCCGACGCTTACCTGAGTACCCGATTC 7240
480 aPheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheA 497
7241 ATACACCAATGAACATCTATCATCTGAGGCTATTCATTTCTCTTAA 7290
497 IAserPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
7291 CAGCA...GTAAATATTATTAATTTTCAATGATTTGAGAGCTTGGCTTCG 7337
514 ArgArgGluThr.....ArgProAsnProTrpGlyGluPheAl 536
7338 AAGCGAAGATCTCTAATAGTAGAAGAACCC.....TC 7369
526 aAspThrLeuGluTyrThrLeuProSerProProAlaHisThrPheG 543
7370 CATTAACCTGGAGTAGATATGATGATGCCGCCACCTACACACATTCG 7419
543 Lu 543
7420 AA 7421
seq_name: /SID58/gcgdata/geneseq/geneseqn/NA2000.DAT:AA257203
seq_documentation_block:
ID AA257203 standard; DNA; 16569 BP.
XX
AC AA257203:
XX
XX
XX 30-MAR-2000 (first entry)
XX
XX Human mitochondrial DNA sequence SEQ ID NO:2.
XX
XX Human: mitochondrial DNA; extramitochondrial DNA: mtDNA; exmlDNA;
XX diagnosis: quantification; detection: dysonomia; Alzheimer's disease;
XX Huntington's disease; Parkinson's disease; schizophrenia; stroke;
XX non-insulin dependent diabetes mellitus; mitochondrial encephalopathy;
XX lactic acidosis; myoclonic epilepsy ragged red fibre syndrome;
XX Leber's hereditary optic neuropathy; ds.
XX
XX Homo sapiens.
XX

```

6106 TCATGTAATACCCATCATATACGAGCGCTTTGGCACTGACGTAGTTCCC 6155

117 LeuGlnIleGlyAlaProAspMetAlaIleProArgMetAsnAspLeu 133

6156 CTAAATATGGGTGCCCGGATATGGCTTTCCCGCATAAACAATATAG 6205

133 rPheTrpLeuPheIleAlaGlyThrAlaMetGlyAlaIleAspLeuPheA 150

150 IAlaProGlyGlyAspGlyGlnIleuGlySerGlyValGlyTrpValLeuTyr 166

6206 CTTCGACTCTTACCTCCCTCTCTCCACTCTGCTGCTGCTGCTATAG 6255

6256 TG.....GAGCGCGGAGCAGGAAAGAGGTGTAACAAGCTAC 6290

167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181

6291 CCTCTTTTACGAGGAACTACCTCCACCCCTGGAGCGCTCCGTAAGCTAAC 6340

181 alIlePheAlaValHisLeuSerGlyValAspSerIleMetGlyAlaIleA 198

6341 CATCTCTCCCTTACCCCTACAGGTGTCTCTCTATCTTACGGCCATCA 6390

198 smMetIleThrThrPheLeuAsnMetArgAlaIleProGlyMetThrLeuHis 214

6391 ATTTCATCACACATATATCAATATATAAACCCCTCCCATACCCCATAC 6440

215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIleLe 231

6441 CAACAGCCCCCTTCGTGCTGATCCGCTCTTAATACAGAGCTCTCTCTCT 6490

231 uIleAlaIleuProValIleuAlaGlyAlaIleThrMetLeuLeuThrAspA 248

6491 CCTATCTCTCCAGCTGCTGATGCTGCGGCTATCATATCTATCTACAGAC 6540

248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 264

6541 GCAACCTCCACACACCCTCTTCGACCCCGGAGAGAGACCCCAT 6590

265 LeuTyrClnHisIleLeuTrpPhePheGlyHisProGlyValTyrIleAl 281

6591 CTATACCAACACCATATCGATATTTTGGCTACCCCTGAAGTTATATCT 6640

281 eIleLeuProGlyPheGlyIleIleSerHisValValSerThrPheSer. 297

6641 TATCTCTACAGGCTTCGGATATCTCCCATATCTTACCTTACCTATCTCC 6690

298 ..LysIleProValPheGlyTyrIleuProMetValTyrAlaMetValAla 313

6691 GAAAAAAGAACCATTTGGATACATAGCTATGGTCTGAGCTATGTATATCA 6740

314 IlegIValIleuGlyPheValValTrpAlaHisHisMetCysThrValGl 330

6741 ATTTGGCTTCTGGGTTATGCTGTGAGCACACACCATATATTATACAGTAG 6790

330 yMetSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleA 347

6791 AATGAGAGTGAACACACAGACATATTACCTCCGCTACCATATATCATCG 6840

347 lAlaIProThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGly 363

6841 CTATCCCGACCGCGCTCAAAAGATATTTAGCTGACTCGCCACACTCCAGCA 6890

364 GlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetPh 380

6891 AGCAATATGAAGATATCTGTGACAGTCTCTGAGCCCTAGACATTCATCT 6940

380 eLeuPheThrValGlyGlyValThrGlyIleValIleuAlaGlnAlaGlyL 397

6941 TCTTTACACCGTAGGTGCGCTGACATGTGATATAGCAAACTCATCAC 6990

397 euAspArgLysIleHisAspThrTyrTyrValValAlaHisPheHisTyr 413

6991 TAAACATCTGCTACTACGACACAGTACTACGTTGTAGCCCACTTCCAGTAT 7040

```

414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaIleGlyIleTyrPheTy 430
7041 GTCTTATCATAGAGACTGATTTGCGATCAAGAGAGGCTTCATTTCACTG 7090
430 rMetProLysPheSerGlyAlaGalaPheProGlyIleTyrPheAlaLysLeuH 447
7091 ATTTCCCTATTTCTCGAGCTACACCTAGACCACTAGCCCAAAATCC 7140
447 IsPheThrPhePheIleGlyAlaAsnValIlePhePheProGlnHis 463
7141 ATTTCACTATTCATTTTCATCGCGCTAAATCACTTTCTTCCCAACAAC 7190
464 PheLeuGlyArgGlnGlyMetProArgArgTyrIleAspTyrProGlnI 480
7191 TTTTCGGCTATCCGGAATGCCCGGACGTTACTCGGACTACCCGAGTC 7240
480 aPheAlaLeuTyrPheAsnLysValSerSerTyrGlyAlaPheLeuAlaPheA 497
7241 ATACACCATGAAACATCCTATCATCTGTAGGCTCATTTCACTTCTCAA 7290
497 IsSerPheLeuPhePheIleValIlePheValIleTyrThrLeuValAlaGly 513
7291 CAGCA...GTAAATATTAATAATTTTCATGATTTGAGAACCTTCGCTTCG 7337
514 ArgArgGluThr.....ArgProAsnProTyrPheGlyIlePheAl 526
7338 AAGCGAAAGTCTCTATAGTAGAAGACCC.....TC 7369
526 aAspThrLeuGlyIleTyrPheLeuProSerProProAlaHisThrPheG 543
7370 CATTAACCTGAGAGTGCATATGATGATGCCCGCCACCTACCACATTCG 7419
543 Lu 543
7420 AA 7421

seq.name: /SID58/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV43718
seq_documentation_block:
ID AAV43718 standard; cDNA to mRNA; 1539 BP.
XX
AC AAV43718;
XX
DT 16-NOV-1998 (first entry)
XX
DE Cancer associated gene fragment 3.
XX
KW ds; cancer; PCR: Northern blotting; ribonuclease protection assay;
XX
diagnosis; metastatic cancer.
XX
OS Synthetic.
XX
PN WO9837187-A1.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98WO-JP00667.
XX
PR 21-FEB-1997; 97JP-0052508.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Asada K, Hino F, Kato I, Mukai H, Yoshikawa Y;
XX
WPI: 1998-467552/40.
XX
PT Detection of cancer cells in tissue samples - by changes in mRNA
PT expression compared to normal tissue of specific cancer-associated
PT gene sequences
XX
PS Claim 1; Page 43-44; 92pp; Japanese.
XX

```

```

CC The cancer associated gene fragments AAV3716-V43731 can be used to
CC detect cancer cells in tissue samples or biological fluids. They are
CC normal tissue of one or more cancer-associated genes whose cDNA
CC stringently hybridises to the nucleic acid fragments. The change in
CC expression may be an increase or a decrease compared to normal tissue.
CC The mRNA expression may be determined by PCR, Northern blotting or
CC ribonuclease protection assay, or by determining the change in the amount
CC of protein encoded by the gene(s) as compared to normal tissue, for
CC example by using a labelled antibody recognising the protein. Detection
CC of cancer cells in tissues other than the primary tumour site.
CC
XX
SO Sequence 1539 BP; 417 A; 463 C; 249 G; 410 T; 0 other;

alignment_scores:
      Quality: 1325.00      Length: 534
      Ratio: 3.193      Gaps: 7
Percent Similarity: 77.715      Percent Identity: 50.187

alignment_block:
US-09-712-768-2 x AAV43718 ..

Align seg 1/1 to: AAV43718 from: 1 to: 1539

17 PhePheThrArgTyrPheMetSerThrAsnHisLysAspIleGlyLeu 33
   ||| .....|||
4  TTGGCGGACCGGTGAGCTATTTCTACAAACCAAGACATTTGGAACT 53
   |||||.....:
33 uTyrLeuValAlaAlaGlyValAlaGlyPheIleSerValLeuPheThrV 50
   |||||.....:
54 ATACCTATTATTCGGCCGATGAGCTGAGTGCAGGACAGCTTAACCC 103
   ::::|.....|
50 aTyrMetArgLeuGlyLeuMetAspProGlyValGlnTyrMetCysLeu 66
   ::::|.....|
104 TCCTTATTGCGAGCGGACCTGGCCAGCCAGCC..... 135
67 GluGlyAlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnG 83
   ::::|.....|
136 .....|
83 yHisLeuTyrPheAsnValMetValThrTyrHisGlyIleLeuMetPheP 100
   |||||.....|
153 CCACATCTACAAAGCTTATGCTACAGCCCATGCAATTGTATATATCTCT 202
100 heValGlyIleProAlaLeuPheGlyIlePheGlyAsnTyrLeuMetPro 116
   |||||.....|
203 TCATGATTAATACCATCAATTCGAGGCTTTGGCAACTGACAGTTCC 252
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeu 133
   ||| .....|
253 CTAAATAATCGGTGCGCCGATATGCGCTTCCCGCATTAACACATTAAG 302
133 rPheTyrPhePheIleAlaGlyThrAlaMetGlyValAlaIleSerLeuPheA 150
   ||| .....|
303 CTTTGACTCTTACCTCCCTCTCTCTACTCCGCTCGCATGTGCTATAG 352
150 laProGlyGlyAspGlyGlnLeuGlySerGlyValGlyTyrValLeuTyr 166
   ::::|.....|
353 TA.....GAGCGCGGAGCGAGAACAGGTTGAACAGCTCTAC 387
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181
   |||||.....|
388 CTCCTCTTACGAGGAGACTTACCCACCTGAGCGCTCCGTAGACCTAAC 437
181 aIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIleA 198
   ::|||.....|
438 CATCTTCTCTTACACTTACGAGGCTCTCTCTATCTTATGAGGCGCATCA 487
198 smMetIleThrPheLeuAsnMetArgAlaProGlyMetThrLeuHis 214
   |||||.....|
488 ATTTCAATCAACAATTAATTAATAAACCCCTGCAATTAACCCCAATAC 537

```



```

215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpIleIle 231
      ::::::::::::::::::::
538 CAAGCGCCCTTCTGCTGATCGCTCTATTCACAGACGCTCTACTCTT 587
      ::::::::::::::::::::
231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAsp 248
      ::::::::::::::::::::
588 CCAATCTCTCCAGTCCTAGCTGCTGACATCTACTACTACTACTACAGACC 637
      ::::::::::::::::::::
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 264
      ::::::::::::::::::::
638 GCACCTCAACACACCTCTCTGACACCCCGGAGAGGAGACCCCACTT 687
      ::::::::::::::::::::
265 LeuTyrGlnHisIleLeuTrpPhePheGlyHisProGluValTyrIleI 281
      ::::::::::::::::::::
688 CTATACCAACACCTATCTGATTTTGGTACACCTCGAAGTTATATCT 737
      ::::::::::::::::::::
281 eIleLeuProGlyPheGlyIleIleSerHisValValSerThrPheSer. 297
      ::::::::::::::::::::
738 TATCTACCAAGCTTCCGATATCTCCATATTTGTAATCTACTACTCG 787
      ::::::::::::::::::::
298 ..LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
      ::::::::::::::::::::
788 GAAAAAAGAACCAATTGGATACATAGCTATGCTGAGCTATGATATCA 837
      ::::::::::::::::::::
314 ILeuGlyValLeuGlyPheValValTyrPAlaHisHisMetTyrThrValG 330
      ::::::::::::::::::::
838 ATTTGGCTTCTAGGGTTATCTGTCGACACACCATATATTTACAGTAGG 887
      ::::::::::::::::::::
330 YMetSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIle 347
      ::::::::::::::::::::
888 AATGAGACGTAGACACACGACATATTTCACTCCGCTACCATATCATCG 937
      ::::::::::::::::::::
347 lAlaValProThrGlyIleLysIlePheSerTrpIleAlaThrMetTyrGly 363
      ::::::::::::::::::::
938 CTATCCCAACGGCGTCAAGATATTAGCTGACTCCACACTCAGCGA 987
      ::::::::::::::::::::
364 GlySerValGluPheLysSerProMetLeuTrpAlaPheGlyPheMetCph 380
      ::::::::::::::::::::
988 AGCAATATGAATGATCTGCTGCAGCTCTGAGCCCTGAGATTCATCTT 1037
      ::::::::::::::::::::
380 eLeuPheThrValGlyGlyValThrGlyIleValIleuAlaGlnAlaGlyL 397
      ::::::::::::::::::::
1038 TCTTTTCACTAGTGGCTGCTGACTGCACTGATTTAGCAACTCTCC 1087
      ::::::::::::::::::::
397 euAspAlaAlaTyrHisAspThrTyrTyrValValAlaHisPheHisTyr 413
      ::::::::::::::::::::
1088 TAGACATCGTACTACACACGACGACTAGCTGAGCTCACTCCACTAT 1137
      ::::::::::::::::::::
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTy 430
      ::::::::::::::::::::
1138 GTCTCATATAGAGCTGTATTTGCCATTCATAGGAGGCTTCATTCCTG 1187
      ::::::::::::::::::::
430 rMetProLysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuH 447
      ::::::::::::::::::::
1188 ATTTCCCTATTTCTCAGCTACACCTAGCAACCTACCCCAAAATTC 1237
      ::::::::::::::::::::
447 lspHeTrpThrPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
      ::::::::::::::::::::
1238 ATTCCGCTATCATATTCATCGGCTAATCTACTTTCTTCCCAACAAAC 1287
      ::::::::::::::::::::
464 PheLeuGlyArgGlnGlyMetProArgArgTyrIleAspTyrProGluAl 480
      ::::::::::::::::::::
1288 TTTCTCGGCTATCGGAATGCCCGCGAGTCTCGACTCGACTACCCGAGTGC 1337
      ::::::::::::::::::::
480 aPheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheA 497
      ::::::::::::::::::::
1338 ATACACCAATGAAATATCTCATCTCTGTAAGCTCATTCATTTCTCTAA 1387
      ::::::::::::::::::::
497 lAserPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
      ::::::::::::::::::::
1388 CAGCA...GTATATTTATATTTTTCATGATTTGAGAACCTTGGCTTGC 1434

```

```

514 ArgArgGluThr.....ArgProAsnProTrpGlyGluPheAl 526
      ::::::::::::::::::::
1435 AAGCGAAAGTCTCTATAGTAGAGAACCC.....TC 1466
      ::::::::::::::::::::
526 aAspThrLeuGlnTrpThrLeuProSerProProAlaHisThrPheG 543
      ::::::::::::::::::::
1467 CATTAACCTGAGTGAATATATGATGCCCCCACCCTACACACATCTCG 1516
      ::::::::::::::::::::
543 lu 543
      ::::::::::::::::::::
1517 AA 1518

```

```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF68879
seq_documentation_block:
ID AAF68879 standard; DNA: 1735 BP.
AC AAF68879:

```

```

XX 12-APR-2001 (first entry)
DE Mitochondrial cytochrome c oxidase subunit I coding sequence.
XX Mitochondria; cytochrome C oxidase; COX; Alzheimer's disease; ss.
XX Homo sapiens.
XX OS
XX US6171659-BL.
XX 09-JAN-2001.
XX 30-MAR-1995; 95US-0413740.
XX 30-MAR-1994; 94US-0219842.
XX (MITO-) MITOKOR.
XX Herrstadt C. Parker WD;
XX WPI: 2001-136875/14.
XX Targeting conjugate molecule to mitochondria having defective
XX cytochrome C oxidase activity for diagnosing Alzheimer's disease,
XX involves contacting mitochondria with a conjugate of targeting molecule
XX and toxin -
XX Disclosure: Fig 1; 88pp; English.
XX The present invention relates to a method for selectively accumulating
XX a mitochondrial disabling or destructive amount of a conjugate molecule
XX in mitochondria having defective cytochrome C oxidase (COX) activity or
XX displaying increased membrane potential. The method involves contacting
XX mitochondria with a conjugate molecule comprising a targeting molecule
XX conjugated to a toxin, where the conjugate or targeting molecule selected
XX accumulates in the mitochondria. The method is useful for diagnosis of
XX Alzheimer's disease (AD), especially sporadic AD. The present sequence
XX is a COX coding sequence.
XX Sequence 1735 BP; 484 A; 515 C; 278 G; 458 T; 0 other;

```

```

alignment_scores:
  Quality: 1311.50      Length: 534
  Ratio: 3.160          Gaps: 8
  Percent Similarity: 77.715      Percent Identity: 50.000

```

```

alignment_block:
US-09-712-768-2 x AAF68879 ..

```

```

Align seg 1/1 to: AAF68879 from: 1 to: 1735
17 PhePheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeu 33
      ||| ||| ::::::::::::::::::::

```

```

75 TTCGCCGACCGTGGACTATCTCTACAAACCAACAGACATTGGACACT 124
33 uTfLeuValAlaIaGlyValAlaGlyPheIleSerValLeuPheThr 50
125 ATACCTATTATTGGCGCATGAGCTGAGTCTCTAGCCACACCTTAAGCC 174
50 aLlyrMetAlaGlyLeuMetAlaPheProGlyValGlnTyrMetCysLeu 66
175 TCCTTATTGAGCGGAGCTGGCGCAGCAGCC..... 206
67 GluGlyAlaArgLeuIleAlaPalasSerGlnThrCysThrAlaSnG 83
207 ..... 111
83 yHISLeuTrpAsnValMetValThrTyrHISGlyIleLeuMetMetPhe 100
224 CCACATCTACACAGCTTATGTCACAGCCCATGATTTGTAAATAACTCT 273
100 heValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 116
274 TCATAGTATATCCATCATATGCGAGGCTTGGCAACGACTAGTCTCC 323
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeu 133
324 CTATTAATCGGTGCCCCGATATGCGCTTCCCGCATTAACACATAG 373
133 rPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPhe 150
374 CTCTGACTCTTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423
150 laPArgGlyAspGlyGlnLeuGlySerGlyValGlyTyrValLeuTyr 166
424 TG.....GAGCCCGAGCAGCAGACAGGTTGACACAGCTTAC 458
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAlaPhe 181
459 CCTCCCTTACGAGGAGACTCTCCACCCCTGACCTCCGTAACCTAAC 508
181 aLlePheAlaValHISLeuSerGlyAlaSerSerIleMetGlyAlaIle 198
509 CATCTTCTCTTACACCTAGCAGGCTGCTCTCTATCTTAGGGGCATCA 558
198 sMeIleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHIS 214
559 ATTTCATACCAACAATTATCATATAAACCCCTGCCATTAACCCAAATAC 608
215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIle 231
609 CAACGCCCTCTTCGCTGATCCGCTTAATCACAGCAGCTACTTCT 658
231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAsp 248
659 CCTATCTCTCCAGTCTAGCTGCTGATCATCTACTACTAACACAGC 708
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 264
709 GCAACCTCAACACCACTTCTTCCAGCCCGGAGGAGAGACCCCAT 758
265 LeuTyrGlnHISIleLeuThrPhePheGlyHISProGlyValTyrIle 281
759 CTATTAACCAACCTATTTCTGATTTTTCGTCACCCGAACTTATATCT 808
281 eLleLeuProGlyPheGlyIleIleSerHISValSerThrPheSer 297
809 TATCTTACAGGCTTGGAAATATCTCCATATTGTATCTTACTCTCG 858
298 ..LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
859 GAAAAAAGAACCACTTGTGATACATAGTATGCTGAGCTATGATATCA 908
314 lIleGlyValLeuGlyPheValAlaTrpAlaHISMetTyrThrValG 330
909 ATTGATTTCTAGGTTTATCGTGTGACACACCATATATTATTAACAGTAG 958

```

```

330 yMetSerLeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIle 347
959 AATAGACGTAGACACACGACGATATTTCACCTCGCTACCATATATCATG 1008
347 lAValProThrGlyIleLeuSIIlePheSerTrpIleAlaThrMetTyrGly 363
1009 CTATCCCAACCGCGGTCAAACTATTATAGCTGACTGCCACACTCCACGA 1058
364 GlySerValGlnPheLysSerProMetLeuThrAlaPheGlyPheMetP 380
1059 ACCATATGAAATGATCTGCTGACGTCTGACCCCTAGATTCATC.. 1106
380 eLeuPheThrValGlyValThrGlyIleValIleuAlaGlnAlaGlyL 397
1107 .CTTTTACCGGTAGGTGGCTGACTGCTGCTTGTATTAACAACATCATCAC 1155
397 euAspArgAlaTyrHISAspThrTyrTyrValAlaIleHISPheHIS 413
1156 TAGACATCGTACTACAGACAGCTACTAGCTTGTAGGCCACTTCCACTAT 1205
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPhe 430
1206 GTCTCATCATAGAGAGCTGATTTTGCATCATAGGAGGCTTCATGAC 1255
430 rMetProLysPheSerGlyArgAlaPheProGlyTyrTrpAlaAlaLysLeu 447
1256 ATTTCCCTATTCTCAGGCTACGCCCTAGACCAAACTCCGCCAAATCC 1305
447 lAspThrTrpPhePheIleGlyAlaAsnValThrPheProGlnHIS 463
1306 ATTTACATATCATATATCATCGCGCTAAATCTACTTCTTCCCAACAC 1355
464 PheLeuGlyArgGlnGlyMetProArgArgTyrIleAspTyrProGly 480
1356 TTTCTGCGCTATCCGGAATGCCCGACCTTACTCGGACTACCCGATGC 1405
480 aPheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPhe 497
1406 ATACACCAATGAACCTCTATCATCTAGTGCATTTCTCTGTA 1455
497 lAsrPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
1456 CAGCA...GTAAATATTAAATTTTCATGATTTAGAAAGCTTCGCTTCG 1502
514 ArgArgGlnThr.....ArgProAsnProTrpGlyIlePheAl 526
1503 AAGCGAAAGTCTTAATAGTAGAAAGAACCC.....TC 1534
526 aAspThrLeuGlyTyrThrLeuProSerProProAlaHISThrPheG 543
1535 CATTAACCTGAGTAGCTATATGATGATGCCCCCACCCTTACACACATTCG 1584
543 Lu 543
1585 AA 1586

```

seq_name: /SID8/gcgdata/geneseq/geneseqn/M2000.DAT:AA257204

seq_documentation_block:

ID: AA257204 standard; DNA; 6691 BP.

XX AA257204;

DT 30-MAR-2000 (first entry)

DE Human extramitochondrial DNA sequence SEQ ID NO:3.

KW Human; mitochondrial DNA; extramitochondrial DNA; mtDNA; exmtDNA;

KW diagnosis; quantification; detection; dystonia; Alzheimer's disease;

KW Huntington's disease; Parkinson's disease; schizophrenia; stroke;

KW non-insulin dependent diabetes mellitus; mitochondrial encephalopathy;

KW lactic acidosis; myoclonic epilepsy ragged red fibre syndrome;


```

397 euAspAtGAlaTyrHisAspThrTyrTyrValAlaHisPheHisTyr 413
PT      ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      TAGACATGCTACTACAGACGACTACTACTACTACTACTACTACT
PS      3572
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyr 430
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3572 GTCCTATCAATAGACGCTGATTGGCATCATAGAGAGCTCATCAGT
PS      3621
430 MetProLysPheSerGlyArgAlaPheProGlyTyrPheAlaAlaLysLeuH 447
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3622 ATTCCCTATCTCTCAGGCTACACCTAGACCAACCTAGCCAAATCC
PS      3671
447 IspHerPhePhePhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3672 ATTTCCCTATCAATATTCATCGGCTAAATCTAATCTTCTCCACACAC
PS      3721
464 PheLeuGlyArgGlnGlyMetProArgTyrTyrIleAspTyrProGluAla 480
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3722 TTTCTGCGCTATCCGGAATGCCCGACGTTACTGCGACTATCCGATGC
PS      3771
480 aPheAlaLeuThrPasnLysValSerSerTyrGlyAlaPheLeuAlaPhe 497
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3772 ATACACCATGAAATATCTATCTATCTAGGCTCATTTCTCTAA
PS      3821
497 IAspPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3822 CAGCAGTATATATATTTTTCATATTTGAGAAGCCTTGCTCGAAG
PS      3871
514 ArgAlaGlyThrThrArgProAsnProTyrGlyGluPheAlaAspThrLeuGly 530
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3872 CGAAAGTCTAATAGTAGAA.....GAACCTCCATAAACCCTGGA
PS      3912
530 uThrPheLeuProSerProProProAlaHisThrPheGly 543
PT      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      3913 GTGACTATATGATGCCGCCACCTCTACACATTTGGA
PS      3952
seq_name: /STD58/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH67734
seq_documentation_block:
ID      AAH67734 standard; DNA; 1743 BP.
XX
XX      AAH67734;
XX
XX      26-SRP-2001 (first entry)
XX
XX      C glutamicum coding sequence fragment seq ID NO: 2769.
XX
XX      Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX      organic acid synthesis; ds.
XX
XX      Corynebacterium glutamicum.
XX
XX      EP1108790-A2.
XX
XX      20-JUN-2001.
XX
XX      18-DEC-2000; 2000EP-0127688.
XX
XX      16-DEC-1999; 99JP-0377484.
XX      07-APR-2000; 2000JP-0159162.
XX      03-AUG-2000; 2000JP-0280988.
XX
XX      (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX      Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX      Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX      WPI; 2001-376931/40.
XX      P-PSDB: AAG92515.
XX
XX      Novel polynucleotides derived from Coryneform bacteria, for identifying
XX      mutation point of a gene, measuring expression of a gene, analysing

```

```

PT      expression profile or pattern of a gene and identifying homologous gene
XX
XX      Claim 8: SEQ ID NO: 2769; 246bp + Sequence Listing; English.
XX
XX      The present invention provides a number of nucleotide and protein
CC      sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC      are useful for identifying the mutation point of a gene derived from a
CC      mutant of coryneform bacterium, measuring expression amount and
CC      analysing the expression profile or expression pattern of a gene derived
CC      from Coryneform bacterium, and identifying a homologue of a gene derived
CC      from coryneform bacterium. Coryneform bacteria are useful for producing
CC      amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC      particularly L-lysine. The present sequence is a nucleic acid described
CC      in the exemplification of the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from the
CC      European Patent Office.
XX
XX      Sequence 1743 BP; 283 A; 541 C; 427 G; 492 T; 0 other;
XX
XX      Alignment scores:
XX      Quality: 1169.50      Length: 556
XX      Ratio: 2.946      Gaps: 11
XX      Percent Similarity: 71.403      Percent Identity: 39.748
XX
XX      Alignment block:
XX      US-09-712-768-2 x AAH67734
XX
XX      Align seg 1/1 to: AAH67734 from: 1 to: 1743
XX
XX      12 HisGluysGlnGlyPhePheThrArgTyrPheMet.....SerThrAs 26
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      55 CATCGACGAGGCG...ACCAAGCATGGTTATGATGACGACGACCGCA 101
XX
XX      26 nHisLysAspIleGlyLeuLeuTyrLeuValAlaIleAlaGlyValAlaGlyP 43
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      102 CCACACACGCTGGCGCATTTGTATCATCATATTATGCTTCACTCTTCT 151
XX
XX      43 heliSerValLeuPheThrValTyrMetArgLeuLeuMetAspPro 59
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      152 TCTCGGGTGGCTTGATGGCCCTGCTTATCCGAGCGAGCTTTTCACCCCT 201
XX
XX      60 GlyValGlnTyrMetCysLeuGluGlyAlaArgLeuIleAlaAspAlaSe 76
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      202 GGTCTGCGAGTTCTG.....
XX
XX      76 rGlnThrCysThrAlaAsnGlyHisLeuThrPasnValMetValThrTyrH 93
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      217 .....TCTAATGAGCAGTTCAACGAGCTGTTCAACCATGC 250
XX
XX      93 IsGlyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGly 109
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      251 ACGGACGTGTCATGCTGCTGCTGTACGGAACCTCAATTGTTGG...GGT 297
XX
XX      110 PheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspMetAlaPhe 126
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      298 TTTTGCTAATCTACGCTGCTGCACTTCAGATCGTGCGCTGACGTAAGCTTT 347
XX
XX      126 eProArgMetAsnAsnLeuSerPheThrPhePheIleAlaGlyThrAlaM 143
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      348 CCCACGTTTGATGCTTTCGGCTTCTGTGATGACACGACGCTGCTGTGTCG 397
XX
XX      143 etGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySer 159
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      398 CGATGCTGACCGGCTTCTGACCCCGGTGT.....GCTGCC 435
XX
XX      160 GlyValGlyTyrPheValTyrProLeuSer.....ThrArgGluAl 174
XX      ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      436 GACTTGGTGGTACCATGTAATCCCATGCTGTGACGCAATTCATCCGCC 485
XX
XX      174 aGlyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAla 191

```

```

486 AGGCTTGGCTGACATGATGATGTCGGTGGCTGCACTGATG 535
191 eSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsnMetArg 207
536 GCTCCGTTGCTCCGCAATTAACATGCTCACCAACATCTGCTCCGCG 585
208 AlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSerIlePhe 224
586 GCACCTGGTATGACCATGTCCTGATGCCATTTTTCACCTGGAGATATCTT 635
224 eIleThrAlaTrpLeuIleLeuAlaLeuProValIleAlaGlyAlaI 241
636 CGTGTTCCTGCTTCTGCTGCTGATCTTCCACCTGCTGCTGCTGCTG 685
241 IeThrMetLeuLeuThrAspArgAsnIleGlyThrPhePheAsnPro 257
686 CACTGGGTCTCTGATGACCGCAAGCTGTGGACACCTGTACGATCCA 735
258 AlaGlyGlyGlyAspProIleLeuTrpGlnHisIleLeuTrpPhePheG 274
736 GCTAACGGCGGCTCC...CTCCTGGGAGACACCTGTTCGTTGCTTCG 782
783 ACACCTGAGTTTACGTTCTGCGGCTCCGCTCTGCGCATCTTCTTCG 832
274 vHisProGlyValTrpIleIleIleLeuProGlyPheGlyIleIleSerH 291
291 IValValSerThrPheSerLysLysProValPheGlyTrpLeuProMet 307
833 AGATCATCTCTGCTGTCCTCCGTAAGCCAAATGTTGCTGTTACGTGGCTG 882
308 ValTrpAlaMetValAlaIleGlyValLeuGlyPheValValTrpAlaH 324
883 ATCTTGCAACCTGTGCCATGTGTGTCACATGCTGATGCTGCTGCTGCA 932
324 shiMetTrpTrpValGlyMetSerLeuThrGlnInsTrpPheMetL 341
933 CCACATGTTCTGTTACTGCGGCAAGTTTGTCT...CCGTTCTTCTCTCT 976
341 eulAlaThrMetValIleAlaValProThrGlyIleLysIlePheSerTrp 357
977 TCATGACGTTCTGATTCGCGTTCCTACCGGCGTTAACTTCAACTGG 1026
358 IleAlaThrMetTrpGlySerValGluPheLysSerProMetLeuTrp 374
1027 GTTGGAACCATGTGGAAGGCTCACATCACTTGGGAAACCCCAATGATCTG 1076
374 palArpGlyPheMetPheLeuPheThrValGlyGlyValThrGlyIleV 391
1077 GTCTGTGGCTTCATGCTACCTTCCCTCGGCGGTGCTGACCGCATTA 1126
391 allLeuAlaGlnAlaGlyLeuAspArgAlaTrpHisAspTrpTrpVal 407
1127 TGCTGGGCTCCCCACCATGAGCTTCACCTGGCTGCTCTCTCTCTCTG 1176
408 ValAlaHisPheHisTrpValMetSerLeuGlyAlaIlePheAlaIleP 424
1177 ATCGGCACTTCACCTACACCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1226
424 eAlaGlyIleTrpPheTrpMetProLysPheSerGlyArgAlaPheProG 441
1227 TGCAGGCGTTTACTCTGTTCCGAMATGATGACCGCGCATGTGACG 1276
441 IuTrpAlaAlaLysLeuHisPheTrpThrPhePheIleGlyAlaAsnVal 457
1277 AGGCTCTTGCAAGATCCACTTCTGTTGACCTTGTGCGGTTTCCACGCA 1326
458 ThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProArgArgTrp 474
1327 ACCTTCTCATCCAGCATGCGGCGCAACATGATGATGCCACGTCGTTA 1376
474 rIleAspTrp.....ProGluAlaPheAlaLeuTrpAsnLysValSer 489

```

```

1377 CGCTGACTACCTGATTCGATGGTTTCACATCTACAAACGATCTCCA 1426
489 eTrpGlyAlaPheIleAlaPheAlaSerPheLeuPhePheIleValIle 505
1427 CCGGTTCTCTCTCTCTCTCTGCTGGCTGCTGCTGCTGCTGCTGCTG 1476
506 PheValTrpThrLeuValAlaGlyArgArgGluThrArgProAsnProTr 522
1477 GTCCTCAAGTCTGCGCCCTACGCTGACGCTTACCGTTGATGATTCCTG 1526
522 PGIleGluPheAlaAspThrLeuGluTrpThrLeuProSerProProProA 539
1527 GGGT...TACGGCAATCCCTGGAGTGGCAACCTGCGCTCTCTCTC 1573
539 IahisThrPheGluThrLeuProLys.....ArgSerAspTrp 551
1574 GCCACAACTTCGATCCCTGCTGATCCGCTCCGACGCCCTGCGTTC 1623
552 AspLysHisProSerHis 557
1624 GAGCTGCACTACCGGCAC 1641
seq_name: /STD8/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH68532
seq_documentation_block:
ID AAH68532 standard; DNA: 349980 BP.
AC AAH68532:
DE 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX Corynebacterium glutamicum.
XX EP1108790-A2.
XX 20-JUN-2001.
XX 18-DEC-2000; 2000EP-0127688.
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Disclosure; SEQ ID NO: 7067; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of corynebacterium bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Corynebacterium bacterium, and identifying a homologue of a gene derived
XX from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed

```

CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

Sequence 349980 BP, 80900 A, 98397 C, 92139 G, 78544 T, 0 other;

alignment_scores:

Quality: 1169.50 Length: 556
 Ratio: 2.946 Gaps: 11
 Percent Similarity: 71.403 Percent Identity: 39.748

alignment_block:

US-09-712-768-2 x AAH68532/rev ..

Align seg 1/1 to reverse of: AAH68532 from: 1 to: 349980

12 HlaGluLysGlnGlyPhePheThrArgTrpPheMet.....SerThrAs 26

272751 CATGCACGCAAGGCG...AGCAAGCATGTATATGATGACACACACCGA 272705

26 nHISLysAspIleGlyLeuLeuTyrLeuValAlaIleGlyValAlaGly 43

272704 CCACAGACGCTGGGCGATATGATCATCATATATGCTTCAGCTTCTCT 272655

43 heIleSerValLeuPheThrValTyrMetArgLeuGluLeuMetAspPro 59

272654 TCCTGGGTGCTGTATGCGCCCTGCTTATCCGAGCGGAGCTTTTCACCCCT 272605

60 GlyValAlaGlnTyrMetCysLeuGluGlyAlaArgLeuIleAlaAspAlase 76

272604 GGTTCGACAGTCTCTCT..... 272590

76 rGlnThrCysThrAlaAsnGlyHisLeuTrpAsnValMetValThrTyrH 93

272589TCTATGACAGCTTCACACGCTGTGACACCAATGC 272556

93 IsglyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGlyGly 109

272555 ACGGAACGTGATGCTGCTGCTGACGAGCTCCCAATTGTTGG...GGT 272509

110 PheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspMetAlaPh 126

272508 TTTCCTACTACGTCTGCTGCCACTTCACATGCGTCCGCTGACCTACTT 272459

126 eProArgMetAsnAsnLeuSerPheTrpLeuPheIleAlaGlyThrAlaM 143

272458 CCCACGTTTGAATGCTTTCGGCTTGGATCACCACCGTGGGGTGTGC 272409

143 eGlyValAlaSerLeuPheAlaProGlyIlyAspGlyLeuGlySer 159

272408 CGATGCGACCGGCTTCCTGACCCGCGGTGT.....GCTGCC 272371

160 GlyValGlyTyrValLeuTyrProPheLeuSer.....ThrArgGluAl 174

272370 GACTTGGGTGGACATGATACCTCCACTGTCTGACGCAATTCACCTCCC 272321

174 aglyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAlas 191

272320 AGGCTTGGCTGTACATGATGATGCTGCGTCCGCTGCACTGATTTG 272271

191 eSerIleMetGlyAlaIleAsnMetIleThrThrPheLeuAsnMetArg 207

272270 GCTCCGCTGCTCCGAATTATACATGCTCACACCATCTCTGCTCCGCG 272221

208 AlaProGlyMetThrLeuHisLysValProLeuPheSerTyrSerIlePh 224

272220 GCACCTGGATATACATGTCTCCGATGCTTATTTTACCTGGATATCTT 272171

224 eIleThrAlaTrpLeuIleLeuLeuAlaLeuProValLeuAlaIlyAla 241

272170 CGTGTCTTCCTGCTTCTGCTGCTGATGATTTCCACCTGCTGCTGCTG 272121

241 IeThrMetLeuLeuThrAspArgAsnPheGlyThrThrPhePheAsnPro 257

272120 CACTGGCTTCTGTATGACCGCAAGCTGTGTGACACCGTATCATCA 272071

258 AlaGlyIleGlyAspProIleLeuTyrGlnHisIleLeuThrPhePheG 274

272070 GCTAACGGCGGCTCC...CTCCGTGGACACACCTCTCTGCTGTGCG 272024

274 YHISProGlyValTyrIleIleIleLeuProGlyPheGlyIleIleSerH 291

272023 ACACCTGAGGTTTACGTTCTGCGCTGCCGCTTCTGCGCATTTCTGCG 271974

291 IValValSerThrPheSerTyrLysProValPheGlyTyrLeuProMet 307

271973 AGATATTCCTGCTTCTCCCGTAACCCATTTCTGATGCTGCGCTG 271924

308 ValTyrAlaMetValAlaIleGlyValLeuGlyPheValIlyAlaI 324

271923 ATCTTGCAACCTGTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271874

324 sHISMetTyrThrValGlyMetSerLeuThrGlnGlnSerTyrPheMet 341

271873 CCACATGTTCTGTTACTGGCGGCGCTTCTCT.....CCGTTCTTCTCT 271830

341 eValIleThrMetValIleAlaValProThrGlyIleLysIlePheSerTyr 357

271829 TCATGACGCTCTGATTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 271780

358 IleAlaThrMetTyrPylGlySerValGlyLeuPheLysSerProMetLeu 374

271779 GTTGGACCATGTGGAGAGGACATCATCTTGGGAAACCCCAATGATCTG 271730

374 PAlaPheGlyPheMetPheLeuPheThrValGlyIlyValThrGlyIle 391

271729 GTCGTGCTTCATAGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 271680

391 aLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyrVal 407

271679 TGTGGCGCTCCGCCACACCTGACTTCCACTTGGCTGCTGCTGCTGCTG 271630

408 ValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAlaIlePh 424

271629 ATGCGACATTCACATACACCTTCTGCTGCTGCTGCTGCTGCTGCTG 271580

424 eAlaGlyIleTyrPheTyrMetProLysPheSerGlyAlaGlyAlaPhePro 441

271579 TGCAGCGCTTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271530

441 IutrrAlaAlaLysLeuHisPheTyrThrPhePheIleGlyAlaAsnVal 457

271529 AGCGTCTTGGCAAGATCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 271480

458 ThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProArgTyr 474

271479 ACCTTCTCATCCAGCACTGGGGGGAACATGAGGTATGACACGCTGTTA 271430

474 rIleAspTyr.....ProGluAlaPheAlaLeuTrpAsnLysValSer 489

271429 CGGTGACTACCTGGATTTGATGATGTTTACCATCATCAACCAAGATTTCCA 271380

489 eTyrGlyValPhePheLeuAlaPheAlaSerPheLeuPheIleValIle 505

271379 CCGTGTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 271330

506 PheValIlyThrLeuValAlaGlyArgArgIlyThrArgProAsnProTr 522

271329 GTCTTCAAGTCTCTGCGCTGACGAGAGCTGTTACCTGATGATCTCTCT 271280

522 pGlyGluPheAlaAspThrLeuGluTyrThrThrLeuProSerProProAla 539

271279 GGGT...TACGGCAACTCTCCGAGTGGGCAACCTCTGCTGCTGCTGCTG 271233

539 IaHisThrPheGlyThrLeuProLys.....ArgSerAspTyr 551

.....

.....

.....

.....

271232 GCCACAACTTCGATCCCTGCTCGATCCGCTCCGAGCCCTGCTTC 271183

552 ASPLYSHIPROSERHIS 557

271182 GAGCTGCCTACTACCCGAC 271165

seq_name: /SID58/gcgdelta/geneseq/geneseqn/NA2001.DAT:AAF71686

seq_documentation_block:

ID AAF71686 standard; DNA; 1779 BP.

AC AAF71686;

30-APR-2001 (first entry)

Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:653.

Corynebacterium glutamicum; carbon metabolism and energy production;
SMP protein; sugar metabolism and oxidative phosphorylation protein;
fine chemical production; organic acid; proteinogenic amino acid;
nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
diagnosis; Corynebacterium diptheriae; evolutionary study; ds.

Corynebacterium glutamicum.

WO200100844-A2.

04-JAN-2001.

23-JUN-2000; 2000MO-1B00943.

25-JUN-1999; 99US-0141031.

08-JUL-1999; 99DE-1031412.

08-JUL-1999; 99DE-1031413.

08-JUL-1999; 99DE-1031419.

08-JUL-1999; 99DE-1031420.

08-JUL-1999; 99DE-1031424.

08-JUL-1999; 99DE-1031428.

08-JUL-1999; 99DE-1031431.

08-JUL-1999; 99DE-1031433.

08-JUL-1999; 99DE-1031434.

08-JUL-1999; 99DE-1031435.

08-JUL-1999; 99DE-1031510.

08-JUL-1999; 99DE-1031562.

08-JUL-1999; 99DE-1031634.

09-JUL-1999; 99DE-1032180.

09-JUL-1999; 99DE-1032227.

09-JUL-1999; 99DE-1032230.

09-JUL-1999; 99US-0143208.

14-JUL-1999; 99DE-1032924.

14-JUL-1999; 99DE-1032973.

14-JUL-1999; 99DE-1033005.

27-AUG-1999; 99DE-1040765.

31-AUG-1999; 99US-0151572.

03-SEP-1999; 99DE-1042076.

03-SEP-1999; 99DE-1042079.

03-SEP-1999; 99DE-1042086.

03-SEP-1999; 99DE-1042087.

03-SEP-1999; 99DE-1042088.

03-SEP-1999; 99DE-1042095.

03-SEP-1999; 99DE-1042123.

03-SEP-1999; 99DE-1042125.

(BADI) BASF AG.

Pompejus M, Kroegeer B, Schroeder H, Zeider O, Haberhauer G;

WPI: 2001-061975/07.

P-PSDB; AAF79569.

New isolated Corynebacterium glutamicum nucleic acid encoding a sugar metabolism and oxidative phosphorylation protein for production or

modulation of production of fine chemicals e.g. amino acids, carbohydrates or enzymes -

Claim 3; Page 1069-1072; 1246pp; English.

AA71360 to AAF71750 encode the Corynebacterium glutamicum sugar metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243 to AAB 79633 which are involved in carbon metabolism and energy production. The C. glutamicum SMP gene can be used in vectors (II) for expression in host cells and production or modulation of production of fine chemicals, such as, an organic acid, a proteinogenic or nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins (II) encoded by them are used for diagnosing the presence or activity of Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells containing them are used to map genomes of organisms related to C. glutamicum, identify and localise C. glutamicum sequences of interest, in evolutionary studies, in determining SMP protein regions required for function, in modulating SMP protein activity, in modulating the metabolism of sugars, and in modulating high-energy molecule production in a cell (i.e. ATP, NADPH).

Sequence 1779 BP; 294 A; 548 C; 435 G; 502 T; 0 other;

alignment_scores:

Quality: 1166.50

Ratio: 2.946

Percent Similarity: 71.223

Percent Identity: 39.748

Align seg 1/1 to: AAF71686 from: 1 to: 1779

12 HisGluTyrGlnGlyPhePheThrArgTyrPheMet.....SerThras 26

68 CATGCACGCCAAGGCG...AGCAAGCATGGTTATGTATGCACCAACCA 114

26 HisLysAspIleGlyLeuLeuTyrLeuValAlaAlaGlyValGlyP 43

115 CCACAGCAGCGCTGGCATATATATATATATATATATATATATATATAT 164

43 helSerValLeuPhePheThrValTyrMetArgLeuGluLeuMetAspPro 59

165 TCCTCGGTGGCTTGTATGGCCCTGCTTATCGAGGAGGAGCTTTACCCCT 214

60 GlyValGlnTyrMetCysLeuGluGlyAlaArgLeuIleAlaAspAla 76

215 GCHTGCAGTTCCTG..... 229

76 rGlnThrCysThrAlaAsnGlyHisLeuTyrPAsnValMetValThrTyr 93

230TCTAATGACAGTTCACACAGCTGTTCACCAATGC 263

93 IsGlyIleLeuMetMetPhePheValGlyIleProAlaLeuPheGly 109

264 ACGGACGTGTATGCTGCTGTACGAGACCTCAATGTTGG...GGT 310

110 PheGlyAsnTyrLeuMetProLeuGlnIleGlyAlaProAspMetAlaPh 126

311 TTTCGTAACGTCCTGCGCAGCTTCAGATCGGTGGCTGACGTAGCTTT 360

126 eProArgMetAsnAsnLeuSerPheThrLeuPheIleAlaGlyThrAla 143

361 CCACGTTGTAATGCTTTCGGCTTGTGATCACACACCGCTGCTGTCG 410

143 etGlyValAlaSerLeuPheAlaProGlyGlyAspGlyGlnLeuGlySer 159

411 CGATGCTACCGCGCTTCGACCCCGGGTGGT.....CTGCGC 448

```

160 GlyValGlyTrpValLeuTyrProLeuSer.....ThrArgGluAl 174
    |||.....|||.....|
449 GACTTCGGTGGACCATGTAATCCCTGTCGACCATTCATGACGCC 498
    |||.....|
174 aGlyTyrSerMetAspLeuAlaIlePheAlaValHisLeuSerGlyAla 191
    |||.....|
499 AGGCTTGGCTGTCGACATGATGTCGGTGTGGTGCMAACGTGATG 548
    |||.....|
191 eSerIleMetGlyAlaIleAsnMetIleThrPheLeuAsnMetArg 207
    |||.....|
549 GCTCGGTGGTCCGCAATTAACATGCTGCACCATCTCTGCTCCGCG 598
    |||.....|
208 AlaProIleMetThrLeuHisLysValProLeuPheSerTrpSerIlePhe 224
    |||.....|
599 GCACCTGGATGACCAATGTCCTATGCTATTTTCACCTGGAAATATCT 648
    |||.....|
224 eIleThrAlaTrpLeuIleLeuAlaLeuProValIleAlaGlyAlaI 241
    |||.....|
649 CGTTGTTTCGCTTCTGCTGCTGATCTCCACAGCTGCTGCTGCTG 698
    |||.....|
241 LeThrMetLeuLeuThrAspArgAsnPheGlyThrThrPheAsnPro 257
    |||.....|
699 CACTGGGTGTTCTGTATGACCGCAACCTGGTGGACACCTGATGATCA 748
    |||.....|
258 AlaGlyGlyAspProIleLeuTyrGlnHisIleLeuTrpPhePheG 274
    |||.....|
749 GCTAACGGGGCTCC...CTCCTGGGACACCTGTTCTGCTGCTGCTG 795
    |||.....|
274 yHisProGluValTyrIleIleIleLeuProGlyPheGlyIleIleSer 291
    |||.....|
796 ACACCTGAGGTTTACGTCTGCGCGCTGCTTCTGCGCATGTTTCTG 845
    |||.....|
291 IsValValSerThrPheSerLysLysProValPheGlyTyrLeuProMet 307
    |||.....|
846 AGATCATCTCTGTTCTCCGTAAGCAATGTCGTTACGTGCGCTG 895
    |||.....|
308 ValTyrAlaMetValAlaIleGlyValLeuGlyPheValIleTrpAla 324
    |||.....|
896 ATCTCGCAACCTGTCATGTCGTCACATGTCATGTCGTCGTCGTC 945
    |||.....|
324 sHisMetTyrThrValGlyMetSerLeuThrGlnGlnSerTyrPheMet 341
    |||.....|
946 CCACATGTTCTGCTGCTGCGGCAAGTTTCTCTT.....CCGTTCTTCTCT 989
    |||.....|
341 euAlaThrMetValIleAlaValProThrGlyIleLysIlePheSerTr 357
    |||.....|
990 TCATGACCGTCTCTGATTCGCTGCTCTACCGCGCTTAAGTCTCTCACTG 1039
    |||.....|
358 IleAlaIleMetTrpGlyLysValGlnPheLysSerProMetLeuTr 374
    |||.....|
1040 GTTGGACACCTGTCGACGTCACATGTCGTAACCCCATGATGTCG 1089
    |||.....|
374 PAlaPheGlyPheMetPheLeuPheThrValGlyValThrGlyLys 391
    |||.....|
1090 GTCTGTGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
    |||.....|
391 aLeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspTrpTyrTyrVal 407
    |||.....|
1140 TGCTGGCTGCTCCSCACCATGATGTCACATGGCTGACCTGCTGCTG 1189
    |||.....|
408 ValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAlaIle 424
    |||.....|
1190 ATGCGGCACTTCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1239
    |||.....|
424 eAlaGlyIleTyrPheTyrMetProLysPheSerGlyAlaGlyAlaPhe 441
    |||.....|
1240 TGCAGGGGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1289
    |||.....|
441 LuTrpAlaAlaLysLeuHisPheTyrThrPhePheIleGlyAlaAsnVal 457
    |||.....|
1290 AGCGTCTTGGCAAGATCCACTCTGCTGCTGCTGCTGCTGCTGCTG 1339
    |||.....|
458 ThrPheProGlnHisPheLeuGlyArgGlnGlyMetProArgArgTyr 474
    |||.....|

```

```

1340 ACCCTTCCTCATCCAGCACCTGGTGGCAACATGATGATGCAGCTGTGA 1389
    |||.....|
474 rIleAspTyr.....ProGluAlaPheAlaLeuTrpAsnLysValSer 489
    |||.....|
1390 CGCGACTTACTGCTGATTCGATGTTGATCCATCTCAACACAGATCTCCA 1439
    |||.....|
489 eTyrGlyAlaPheLeuAlaPheAlaSerPheLeuPhePheIleVal 505
    |||.....|
1440 CCGTGTCTACTCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1489
    |||.....|
506 PheValTyrThrIleValAlaGlyArgArgGlyThrArgProAsnProTr 522
    |||.....|
1490 GTCTTCAAGTCTCGGCGCTACGGTGAAGCTGTTACCTGATGATGATCT 1539
    |||.....|
522 PGIyGluPheAlaAspThrLeuGluTrpThrLeuProSerProProPhe 539
    |||.....|
1540 GGGT...TAGGCAACTCCCTGAGTGGGCAACCTTCCTGCTGCTGCTG 1586
    |||.....|
539 IaHisThrPheGluThrLeuProLys.....ArgSerAspTrp 551
    |||.....|
1587 GCCCAACTTCGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1636
    |||.....|
552 AsPlyHisProSerHis 557
    |||.....|
1637 GAGCTGCATACCCGAC 1654
    |||.....|

seq_name: /SID58/gcgdata/geneseq/geneseq/NA2001.DAT:AAH67733
seq_documentation_block:
ID AAH67733 standard; DNA; 1596 BP.
XX
AC AAH67733;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2768.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX
OS organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakajawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
XX
DR P-PSDB: AAG92514.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
PS Claim 8; SEQ ID NO: 2768; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived

```


from Coryneform bacterium, and identifying a homologue of a gene derived from Coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

XX Sequence 1596 BP: 469 A; 384 C; 492 G; 251 T; 0 other;

alignment_scores:
Quality: 1163.00 Length: 542
Ratio: 2.982 Gaps: 9
Percent Similarity: 71.956 Percent Identity: 40.037

alignment_block:
US-09-712-768-2 x AAH67733/rev ..

Align seg 1/1 to reverse of: AAH67733 from: 1 to: 1596

```

24 Serthrshnhtlyslaspiilegyleuleuylrleuvalaialaalya 40
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1567 ACCACCCACCCACGACGCTGGCATATGATCATATATGCTCTTCA 1538
40 lvalglypheileservalrleuphetrvalrleuvalrleuvalrleu 57
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1537 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1488
57 etaspproglrvalrleuylrleuylrleuylrleuylrleuylrleu 73
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1487 TCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1465
74 Aspalaserlntncysrthrvalrleuylrleuylrleuylrleuylr 90
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1464 .....TCTATAGACAGCTTCAACGAGCTT 1439
90 lthrtyrhslgylleuleumetphepvalrleuylrleuvalrleu 107
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1438 CACCATGACGACGACGACGACGACGACGACGACGACGACGACGAC 1389
107 heglylrypheglrlyslrleuylrleuylrleuylrleuylrleu 123
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1388 GG...GGTTTGTACTACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1342
124 MetAlaPheProArGmetAsnleuSerPheTrpLeuPheIleAla 140
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1341 GTAGCTTTCCACGCTTGAATGCTTGGCTTCTGATCACCACCGCTGG 1292
140 ythrAlaMetGlyValAlaSerleuPheAlaProGlyGlyAspGlyGln 157
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1291 TGGTGTGCGATGCTGACCGCTTCTGACCCCGGGTGGT..... 1252
157 euglyserglyValGlyTrpValrleuylrleuylrleuylrleuylr 171
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1251 ..GCTGCGACTTGGCTTGGACCATGACTCCACCTGCTGACGACAT 1204
172 ArgGluAlaGlyTrpSerMetAspLeuAlaIlePheAlaValHisLeu 188
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1203 CACGCCACGCGCTGGCTGCTGATGATGATGATGATGATGATGATGAT 1154
188 rglYAlaSerSerIleMetGlyAlaIleAsnMetIleThrTrpPheLeu 205
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1153 TGGTATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104
205 smetAlaArgAlaProGlyMetThrLeuHisValrleuPheSerTrp 221
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1103 GGCTCGCGACCTGATGACCATGATGATGATGATGATGATGATGATGAT 1054
222 SerIlePheIleThrAlaTrpLeuLeuLeuAlaLeuProValLeuAl 238
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1053 AATATCTTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004

```

```

238 aglyAlaIleThrMetLeuLeuThrAspArgAsnPhelGlyThrThrPhe 255
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1003 CGGTGTGACGCTGGCTGCTGTATGACCGACGCTGCTGCTGCTGCTGCT 954
255 heAsnProAlaGlyGlyAspProIleLeuTrpGlnHisIleLeuTrp 271
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
953 ACATCCACCTAACGCGGCTCC...CTCTGTGACACCTGCTGCTGCTG 907
272 PhePheGlyHisProGlyValrleuylrleuylrleuylrleuylrleu 288
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
906 TTCTTGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857
288 eileSerHisValrleuylrleuylrleuylrleuylrleuylrleu 305
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
856 TGTTCATGATCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
305 euprometValrleuylrleuylrleuylrleuylrleuylrleuyl 321
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
806 TCGGCTGATCTTGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 757
322 TrpAlaHisMetTrpThrValrleuylrleuylrleuylrleuylrleu 338
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
756 TGGCTCACACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 713
338 rPheMetLeuAlaThrMetValrleuylrleuylrleuylrleuylrleu 355
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
712 CTCTCTCTCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663
355 heserTrpIleAlaThrMetTrpGlySerValrleuylrleuylrleu 371
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
662 TCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 613
372 MetLeuTrpAlaPheGlyPheMetPheLeuPheThrValrleuylrleu 388
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
612 ATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 563
388 rglYIleValrleuylrleuylrleuylrleuylrleuylrleuylr 405
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
562 CGCATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
405 yThrValrleuylrleuylrleuylrleuylrleuylrleuylrleu 421
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
512 ACTTCCTGATCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 463
422 AlaIlePheAlaGlyTrpThrPheTrpMetProLysPheSerGlyAla 438
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
462 GCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 413
438 aPheProGlyTrpAlaAlaLysLeuHisPheTrpThrPheIleGly 455
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
412 GATGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 363
455 laAsnValrleuylrleuylrleuylrleuylrleuylrleuylrleu 471
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
362 TCCACGACGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 313
472 ArgArgTrpIleAspTrp.....ProGluAlaPheAlaLeuTrpAsn 486
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
312 CGTGTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 263
486 sValSerSerTrpGlyAlaPheLeuAlaPheAlaSerPheLeuPhe 503
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
262 GATCTCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 213
503 leValIlePheValrleuylrleuylrleuylrleuylrleuylrleu 519
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
212 TCTGACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 163
520 AsnProTrpGlyGluPheAlaAspThrLeuGlyTrpThrLeuProSer 536
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
162 GATCTTGGGCT...TACGCAACCTCTGATGCTGCTGCTGCTGCTGCT 116

```

536 OProProAlaHisThrPheGluThrLeuProLys.....Args 549
 115 TCTCTCTCGGCACACACTTCGCATCTTCCTCTGTATCCGCTCCGAGCC 66
 549 eAspTrpAspLysHisProSerHis 557
 65 CTGCGTTTCGAGTCGACATCCGCCAC 40

seq_name: /SID8/gcdata/geneseq/geneseq/NA2001.DAT:AAH52833

seq_documentation_block:

ID AAH52833 standard; DNA; 1947 BP.

AC AAH52833:

DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1059.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.

OS Staphylococcus epidermidis.

PN MO20134809-A2.

PD 17-MAY-2001.

PE 09-NOV-2000; 2000MO-US30782.

PR 09-NOV-1999; 9905-0164258.

PA (GLAX) GLAXO GROUP LTD.

PI kimmerly wf;

DR WPI: 2001-316495/33.

DR P-PsDB: AAG81983.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis.
 PS Claim 8; Page 308-309; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 1947 BP; 538 A; 367 C; 361 G; 681 T; 0 other;

alignment_scores:

Quality: 1078.50 Length: 550
 Ratio: 2.794 Gaps: 8
 Percent Similarity: 70.182 Percent Identity: 38.545

alignment_block:
 US-09-712-768-2 x AAH52833 ..

Align seg 1/1 to: AAH52833 from: 1 to: 1947

18 PheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeuLeu 34
 88 TATAAGAGATGTTCACATCCGTAGACCAAAAAMAAATCGTATCATGTA 137
 34 rLeuValAlaAlaGlyValAlaGlyPheIleSerValLeuPheThrValT 51
 138 TTTAATTTCTGCCGTATTAACTTTCCTGCTGCTGATCATCGTCTTAA 187
 51 yrMetArgLeuGluLeuMetLaspProGlyValGlnTyrMetCysLeu 67
 188 TGTTACGTACTCAATTACAAATTCGAGTACAAATCTTCG..... 228
 68 GlyAlaArgLeuIleAlaAspLaseGlnThrCysThrAlaAsnGlyH 84
 229GAGCAAA 236
 84 sleuTrpAsnValMetValThrTyrHisGlyIleLeuMetLepPhe 101
 237 CCACGTATTAATGAGATATTACTACGCGCGGTAAATATGATATTATTA 286
 101 aIGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 117
 287 TGGCTATTCGCGTATTATCTTT...GGTTATGCAATGTTGTTATTCATTA 333
 118 GlnIleGlyAlaProAspMetAlaPheProArgMetAsnLaseSerP 134
 334 CAACCTGGTCACGCGATGTCCTCCGTAAATGATACCTAGTATT 383
 134 eTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPheAla 151
 384 CTGGCTATTCTCGCTGATGATTTATTCACACCTATCA...TTTATTG 430
 151 roGlyGlyAspGlyGlnLeuGlySerGlyValGlyTyrValLeuTyr 167
 431 TAGTGGA.....TCACACGCTGCTGGTACCTACACGCA 468
 168 ProLeuSerThrArg...GlnAlaGlyTyrSerMetLaspLeuAlaIle 183
 469 CCACCTGGTGGTGAATTCAGTCCAGGTCACGAGTCACTATATTATTA 518
 183 eAlaValHisLeuSerGlyLaseSerIleMetGlyAlaIleAsnMet 200
 519 TGCATTCGAATATCTGGTATCGGATCGTTAATGACTGGTATCACTCT 568
 200 lethrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHisLysVal 216
 569 TTGTACGATTTAAGATGTAACCCACCAATGAATTTATGCAATGC 618
 217 ProLeuSerThrPheIlePheIleThrAlaTrpLeuIleLeuAla 233
 619 CCAATGTTTCAGTGAACAACATTCATTAACACATTAACGTTATATTAGC 668
 233 aluProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArg 250
 669 ATCCACAGTTCACGTAGACCTGCTTATATGACGTGATGAATTT 718
 250 heGlyThrThrPhePheAsnProAlaGlyGlyAspProIleLeuTyr 266
 719 TTGGACGTAGTTCCTACGTAGCAATGCGGTATGCCAATGCTTTGG 768
 267 GlnHisIleLeuTrpPhePheGlyHisProGlyValTyrIleIleLeu 283
 769 GCAACTCTTCTGGGTATGGGGCCACCTGAAGTTATTCGTATTTT 818
 283 uProGlyPheGlyIleIleSerHisValValSerThrPheSerLysLysP 300
 819 GCCACATTCGTATGACTCAGAAATCATCCCTACTTTGCCCGTAAC 868

300 rovalPheGlyTyrLeuProMetValTyrAlaMetValAlaIleGlyVal 316
:::|||||::: |||::: |||::: |||:::
869 GTTATTCTGGTCATCAAGTATGATTGGCACTGCAGGATCGCATTC 918
317 LeuGlyPheValValTyrAlaHisIleMetTyrThrValGlyMetSerLe 333
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
919 TTAGGTTCTTAGTTGGGTTCCACATTCTTCACATAGGTAATGGTGC 968
333 uThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValProT 350
::: |||::: |||::: |||::: |||::: |||::: |||:::
969 GTTAATTACTCATTTCTCTCATCTCAACAATGTTAATCGGTGCCAA 1018
350 hrGlyIleuIlePheSerTyrPheAlaThrMetTyrGlyIleSerVal 366
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1019 CCGAGGTAACATAATTAACTGGTGTCCACATTATACAAAGTAGAAT 1068
367 GluPheIleSerProMetLeuThrAlaPheGlyPheMetPheLeuPheTh 383
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1069 ACATTGTGAGTCACCTATGCTATTCATAGCATCATCCCTAACTCTT 1118
383 rValGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAspArg 400
::: |||::: |||::: |||::: |||::: |||::: |||:::
1119 ATTGAGAGGGGTACTGGTGAATGCTGCATGCAATGCAATGCACTATC 1168
400 IatYrHisAspThrTyrTyrValValAlaHisPheHisTyrValMetSer 416
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1169 AATATCACACACTTATTTCTTACTAGCTCAGCTCCATACATGCTT 1218
417 LeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetProLy 433
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1219 ACTGGGTAGATTTGCTGCTAGCTGTTAAATCTTGTGATGCCAA 1268
433 srPheSerGlyArgAlaPheProGluTyrAlaAlaIleuHisPheTyrP 450
::: |||::: |||::: |||::: |||::: |||::: |||:::
1269 AATGATGGCTACAAAGTAAATGAACATTAACAACATGGCTTGTGT 1318
450 hrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeuGly 466
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1319 TCTTCATGATCGATTTACGTTTGTCTTACCAACATTCATTCAGT 1368
467 ArgGlnGlyMetProArgArgTyrIleAspTyrProGluAlaPheAlaLe 483
::: |||::: |||::: |||::: |||::: |||::: |||:::
1369 TTAGATGTGATGCCACGCTGCTATACACTTACGCTTGTGATGCTT 1418
483 uTrp.....AsnIleValSerSerTyrGlyAlaPheLeuAlaPheAla 498
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1419 GTGGTTACTAAGCTTCATCGTACATCGTGCAGTATGATGCAATTG 1468
498 erPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGlyArg 514
::: |||::: |||::: |||::: |||::: |||::: |||:::
1469 GATTCTATTCTTAGTTCAGATGCTTTATATAGCATATCAAGCTCCA 1518
515 ArgGluThrArgProAsnProTyrGlyGluPheAlaAspThrLeuGluTr 531
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1519 CGTGAAGCTACTGGAGATACTGGATGAGCTGGTGTCTTATAGANTG 1568
531 rThrLeuProSerPro...ProProAlaHisThrPheGluThrLeuProL 547
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1569 GTCTACAGCATCAGCTATTCACCTAAATACAACTTGTCTATCAGCTCTG 1618
547 ysArgSerAspTyrAsp.....LysHisProSerHis 557
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
1619 ATTGGAAATGATACATTCATTGTTGATATGAAGAACAATGGTCTCAT 1668

```

1459 CAGCA...GTAATATTAATTAATTTTCATGATTTGAGAGACCTTCGCTTCG 1505
514 ArgArgLgUthr.....ArgProAsnProTrpLgUhpheal 526
      :::::
1506 AAGCGAAAGTCCCTAATAGTAGAAGAACCC.....TC 1537
526 aAspThLeuGluTrpThLeuProSerProProAlaHisThrPheG 543
      :::::
1538 CATAAACCTGGAGTACTATATGATGATGCCCCACACCTTACACACATTCG 1587
543 Lu 543
      ||
1588 AA 1589

seq_name: /c9n2_6/ptodata/2/lna/5B_COMB.seq:US-08-810-599-1

seq_documentation_block:
; Sequence 1, Application US/08810599
; Patent No. 5976798
; GENERAL INFORMATION:
; APPLICANT: PARKER, W. Davis
; APPLICANT: HERRNSTADT, Corinna
; APPLICANT: GHOSH, Soumitra S.
; APPLICANT: FAHY, Eoin
; TITLE OF INVENTION: Methods for Detecting Mitochondrial Mutations
; TITLE OF INVENTION: Diagnostic for Alzheimer's Disease and Methods for Determining
; TITLE OF INVENTION: of Mitochondrial Nucleic Acid
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: US
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/810,599
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/757,438
; FILING DATE: 27 No. 5976798 1996
; APPLICATION NUMBER: US 08/614,072
; FILING DATE: 12 Mar 1996
; APPLICATION NUMBER: US 08/536,036
; FILING DATE: 29 Sep 1995
; APPLICATION NUMBER: US 08/414,969
; FILING DATE: 31 Mar 1995
; APPLICATION NUMBER: US 08/413,740
; FILING DATE: 30 Mar 1995
; APPLICATION NUMBER: US 08/410,658
; FILING DATE: 24 MARCH 1995
; APPLICATION NUMBER: US 08/397,808
; FILING DATE: 3 Mar 1995
; APPLICATION NUMBER: US 08/219,842
; FILING DATE: 30 MARCH 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2105/17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-810-599-1

alignment_scores:
      Quality: 1328.00      Length: 534
      Ratio: 3.192      Gaps: 7
      Percent Similarity: 77.903      Percent Identity: 50.187

alignment_block:
US-09-712-768-2 x US-08-810-599-1

Align seg 1/1 to: US-08-810-599-1 from: 1 to: 1613

17 phepethrargTrpPheMetSerThrAsnHisLysAspIleGlyLeu 33
   |||  |||  :::::|||||  |||  :::::|||||  |||  |||
75 TTGGCCGACCGGTGACTATTTCTTCAACCAACCAAGACATGGAGACT 124
33 uUryLeuValAlaAlaGlyValAlaGlyPheIleSerValLeuPheTrv 50
   |||||  |||  :::::|||||  :::::|||||  :::::
125 ATACTATTATTCGGCGCATGAGCTGAGTCTAGGACACACTTAAGCC 174
50 aUryMetArgLeuGluLeuMetAspProGlyValAlaGlnTyrMetCysLeu 66
   ::  ::|||  |||||  ::|||  |||
175 TCCTATTTCGAGCCGAGCTGGGCCAGCCAGCCAGCC..... 206
67 GluGlyAlaArgLeuIleAlaAspAlaSerGlnThrCysThrAlaAsnG 83
   ::  ::|||
207 .....AACCTTCTAGGTAAAGA 223
83 yHisLeuTrpAsnValMetValThrTrpHisGlyIleLeuMetPheP 100
   |||  :::::|||||  |||  :::::|||||  |||  :::::|||||
224 CCACATCTACAAAGCTATTCGTACACAGCCCATGCAATTGTATTAATCTCT 273
100 heValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 116
   |||||  |||||  :::::|||||  |||  ::|||
274 TCATGTAATACCATCATATTCGAGGCTTGGCACTGACTAGTCTCC 323
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnLeu 133
   |||  |||||  |||||  |||||  |||||  |||||  |||||
324 CTAATATCGGTGCCCGCATATGCGCTTCCCGCATTAACAACATTAAG 373
133 rPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPhe 150
   |||  |||||  ::  ::  ::|||  |||||  |||||
374 CTTCTGACTCTTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423
150 LaproGlyGlyAspGlyGlnLeuGlySerGlyValGlyTyrValLeuTyr 166
   ::  ::  ::|||  |||||  |||||  |||||  |||||
424 TG.....GAGCCGAGCAGAGACAGCTGTAACAGCTTAC 458
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetLeuAla 181
   |||||  |||||  ::  ::  ::|||  |||||  |||||
459 CTTCTCTTACGAGGAGACTACTCCACCTGGAGCTCCGAGCATCTTAC 508
181 aIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIle 198
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
509 CATCTCTCTCTTACACCTAGAGAGTCTCTCTATCTTAGGGGCACTCA 558
198 smMetIleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuHis 214
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
559 ATTTCATCACACAAATATCAATATAAACCCCTGCCATAACCCCAATAC 608
215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIle 231
   :::::|||||  |||||  |||||  |||||  |||||
609 CAAGCGCCCTCTTCTGTCGATCGCTCTATACACAGCACTCTACTCTT 658
231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAsp 248
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
659 CTAATCTCTCCAGTCTCTAGCTCTGCACTACTATACTACTACAGAGCC 708

```



```
seq_name: /cgn2_6/ploadta/2/fna/6B_COMB.seq;us:-09-097-889-2
: seq_documentation_block:
: Sequence 2, Application US/09097889
: Patent No. 6218117
: GENERAL INFORMATION:
: APPLICANT: Herrnstadt, Corrina
: APPLICANT: Ghosh, Soumitra S.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
: TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY AFTER DETECTABLE
: TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/097, 889
: FILING DATE: 15-JUN-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Rosenman Ph.D., Stephen J.
: REGISTRATION NUMBER: 43,058
: REFERENCE/DOCKET NUMBER: 660088.417
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1659 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-097-889-2
alignment_scores:
Quality: 1328.00 Length: 534
Ratio: 3.192 Gaps: 7
Percent Similarity: 77.903 Percent Identity: 50.187
alignment_block:
US-09-712-768-2 x US-09-097-889-2 ..
Align seg 1/1 to: US-09-097-889-2 from: 1 to: 16569
17 phephethargtirphenetsetrhnshhsyssaepilleglyleule 33
||||| | | : ||||| ||||| ||||| ||||| ||||| ||
5907 TTCGCCGCGACCTTGCAATTCTCTACAACCACAAAGCATTTGGACACT 5956
33 uyrileuvalalaciyalaivayalcipheileservalneuphenhyv 50
||||| |: ||||| |: ||||| |: ||||| |: ||||| |: |||||
5957 ATATCTATTATTCGGCCATGAGCCTGGATCCTTAGCACACCTCTAAGCC 6006
50 altyrMetarLeuglLneuMetaspProglyValgIntTyMeCyLeu 66
:: ::||| ||||| :::|||||
6007 TCCTTATTCGAGCGAGCTGCAGCCAGCCAGGC..... 6038
67 GlUGLyAlArGrLeuIleaIsPaLasergInTrCyThrAlaAsngI 83
:: ::|||
AACCTTCTAGGTACGA 6055
```

```

83  yHisLeuTrpAsnValMetValThrTyrHisGlyIleLeuMetMetPheP 100
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6056 CCACATCTACAGCTTATCTGTACAGCCATGCGATTGTGAATATATCTTCT 6105
100 heValGlyIleProAlaLeuPheGlyIlePheGlyAsnTyrLeuMetPro 116
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6106 TCATAGTAATACCATCATATGAGGAGCTTTGGCAACTGACCTAGTTCC 6155
117 LeuGlnIleGlyAlaProAspMetAlaPheProArgMetAsnAsnLeu 133
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6156 CTATATATAGTGTGCCCCCGATATGGCGTTTCCCGCATTAACAACTAG 6205
133 rPheTrpLeuPheIleAlaGlyThrAlaMetGlyAlaLaserLeuPhe 150
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6206 CTTCCTAGCTCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAG 6255
150 IarProGlyIleAspGlyIleLeuGlySerGlyValGlyTyrValLeuTyr 166
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6256 TG.....GAGGCCGAGAGCAAGCAAGTTGAACAGCTTAC 6290
167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6291 CTCTCCCTTAGAGGAGAACTACTCCACCCCTGAGCCCTCGTAGACCTAAC 6340
181 alIlePheAlaValHisLeuSerGlyAlaSerSerIleMetGlyAlaIle 198
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6341 CATCTTCTCTTACACCTAGCAGGCTCTCTCTCTCTCTCTCTCTCTCTA 6390
198 smMetIleThrPheLeuAsnMetArgAlaProGlyMetThrLeuHis 214
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6391 ATTTCATCAGACACATTTATCAATATAAACCCCTGCCATTAACCAATAC 6440
215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIle 231
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6441 CAACGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6490
231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrAsp 248
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6491 CCTATCTCTCCAGTCCCTAGCTGCTGCGCATCTACTACTACTAACACACC 6540
248 rGAsnPheGlyThrThrPhePheAsnProAlaGlyIleGlyAspProIle 264
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6541 GCAACCTTCACACACCTTCTTCGACCCCGGAGAGAGACCCCATTT 6590
265 LeuTyrGlnHisIleLeuTrpPhePheGlyHisProGlyValTyrIle 281
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6591 CTATACCAACACCTATCTGTATCTGTATCTGTATCTGTATCTGTATCT 6640
281 eIleLeuProGlyPheGlyIleIleSerHisValIleSerThrPheSer 297
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6641 TATCTCTACGAGGCTTCGAAATATCTCCCATTTGTACTACTACTCTCCG 6690
298 ..LysLysProValPheGlyTyrLeuProMetValTyrAlaMetValAla 313
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6691 GAAAAAAGAACCATTTGGATACATAGTATGCTTCGACCTATGATATCA 6740
314 IlleIleValLeuGlyPheValIleTrpAlaHisIleMetTyrThrVal 330
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6741 ATTGGCTTCTAGGTTTATCGTGTAGACACACATATTTTACAGTATAG 6790
330 yMetSerLeuThrGlnIleSerTyrPheMetLeuAlaThrMetValIle 347
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6791 AATAGAGCTAGACACAGCATATTTTCACTCCGCTACCATATATCATCG 6840
347 IValProThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGly 363
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6841 CTATCCCAACGCGCTCAAGTATTTAGCTGACCTGCCACACTCCACGGA 6890
364 GlySerValGlnPheLysSerProMetLeuTrpAlaPheGlyPheMetPh 380
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6891 AGCAATATGAAATGATCTGCTCAGTGGCTGTAGCCCTAGATTCATCTT 6940

```

```

380 eLeuPheThrValGlyValIleThrGlyIleValLeuAlaGlnAlaGly 397
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6941 TCTTTTACCGTAGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6990
397 euAspAlaIleTyrHisAspThrTyrTyrValValAlaHisPheHisTyr 413
    |||.....|||.....|||.....|||.....|||.....|||.....|||
6991 TTAACATCTGATCTACACAGACACACTACTACTACTACTACTACTACTAT 7040
414 ValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyr 430
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7041 GTCTCTCATATAGACTGTATTTGCGCATCATATAGGAGGCTTCATTCAG 7090
430 rMetProLysPheSerGlyArgAlaPheProGlyTrpAlaAlaLysLeuH 447
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7091 ATTTCCTATTTCTCAGGCTACACCTGAGCAACCAACCTAGCCAAATCC 7140
447 IspHeTrpPhePheIleGlyAlaAsnValThrPhePheProGlnHis 463
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7141 ATTTCATCTATCATTTTCATGCGGCTAAATCTACTCTTCTCCACAAAC 7190
464 PheLeuGlyArgGlnGlyMetProArgTrpTyrIleAspTyrProGlyAl 480
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7191 TTTCTGCGGCTATCCGGAATGCGCGGAGCTTACTGAGACTACCCGATCC 7240
480 alPheAlaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPhe 497
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7241 ATACACACATGAAACATCTATCATCTGTAGGCTCATCTTCTCTCTAA 7290
497 IAspPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7291 CAGCA...GTAATATTAATTAATTTTCATGATTTGAGAACGCTTCGCTTG 7337
514 ArgArgGluThr.....ArgProAsnProTrpGlyIlePheAl 526
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7338 AAGCAAAAGTCTAATAGTAGAAGAACCC.....TC 7369
526 aAspThrLeuGlnIleTrpPheLeuProSerProProAlaHisThrPheG 543
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7370 CATTAACCTGAGTACTATATATGATGATGATGATGATGATGATGATG 7419
543 Lu 543
    |||.....|||.....|||.....|||.....|||.....|||.....|||
7420 AA 7421
seq_name: /cgn2_6/protodata/2/ina/6B_COMB.seq:US-08-413-740A-1
seq_documentation_block:
; Sequence 1, Application US/08413740A
; Patent No. 6171859
; GENERAL INFORMATION:
; APPLICANT: HERRNSTADT, CORINNA
; APPLICANT: PARKER, WILLIAM D.
; APPLICANT: DAVIS, ROBERT
; APPLICANT: MILLER, SCOTT W.
; TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
; TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
; NUMBER OF SEQUENCES: 206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20036-5405
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,740A
; FILING DATE:

```


1356 TTCTCGGCGCATCCGGAATGCCCGACGTTACTCGGACTACCCGATGC 1405
 480 aPhaLeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPhe 497
 1406 ATACACCATGAACATCTATCTATCTGTAGGCTCATCTTCTCTAA 1455
 497 LaSerPheLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
 1456 CAGCA...GTAATATATATATTTTCATGATTTGAGAAGCCTTCGTCG 1502
 514 ArgArgGluThr.....ArgProAsnProTrpGlyGluPheAl 526
 1503 AAGGAAAGGCTTAATAGTAGAAGAACCC.....TC 1534
 526 aAspThrLeuGluTrpThrLeuProSerProProAlaHisThrPheG 543
 1535 CATAACTGAGTACTATATGATGATGCCGCCACCTACACACATTGC 1584
 543 Lu 543
 1585 AA 1586

seq_name: /cgn2_6/ptodata/2/ina/PCUS_COMB.seq:PCT-US95-04063-1

seq_documentation_block:
 ; Sequence 1, Application PC/TUS9504063
 ; GENERAL INFORMATION:
 ; APPLICANT: HERRNSTADT, CORINNA
 ; APPLICANT: PARKER, WILLIAM D.
 ; APPLICANT: DAVIS, ROBERT W.
 ; APPLICANT: MILLER, SCOTT W.
 ; TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
 ; TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
 ; NUMBER OF SEQUENCES: 206
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kenyon & Kenyon
 ; STREET: 1025 Connecticut Avenue, N.W.
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20036-5405
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04063
 ; FILING DATE: 30-MAR-1995
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bonham, David B.
 ; REGISTRATION NUMBER: 34297
 ; REFERENCE/DOCKET NUMBER: 2105/7
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 429-1776
 ; TELEFAX: (202) 429-0796
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1735 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: other nucleic acid
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; PCT-US95-04063-1

alignment_scores:
 Quality: 1311.50 Length: 534
 Ratio: 3.160 Gaps: 8

Percent Similarity: 77.715 Percent Identity: 50.000

alignment_block:
 US-09-712-768-2 x PCT-US95-04063-1 ..

Align seg 1/1 to: PCT-US95-04063-1 from: 1 to: 1735

17 PhePheThrArgTrpPheMetSerThrAsnHisLysAspIleGlyLeuLeu 33
 75 TTGCGCGACCGTTGACTATTCTCTACAAACACAAAGACATTGGAACCT 124
 33 uTrpLeuValAlaAlaGlyValValGlyPheIleSerValLeuPheTrp 50
 125 ATACCTATTATTCCGCCGATGAGCTGAGACTCCTAGGACAGCTTAAGCC 174
 50 aTyrMetArgLeuGluLeuMetAspProGlyValGlyIleTyrMetCysLeu 66
 175 TCCTTATTCGAGCCGAGCTGGGCGACGAGC..... 206
 67 GluGlyAlaArgLeuIleAlaAspAlaSerGluThrCysThrAlaAsnG 83
 207AACCTTCTAGGTAAACA 223
 83 yHisLeuTrpAsnValMetValThrTyrHisGlyIleLeuMetMetPheP 100
 224 CCACATCTCAACAGTATATGCTCACAGCCCATGTCATTGTAATATCTTCT 273
 100 heValGlyIleProAlaLeuPheGlyGlyPheGlyAsnTyrLeuMetPro 116
 274 TCATAGTAATACCCATCAATATCGAGGCTTGGCAACTGACTATGTTCC 323
 117 LeuGluIleGlyAlaProAspMetAlaPheProArgMetAsnLeuSe 133
 324 CTAATATATGCGTGGCCCGATATGCGCTTCCCGCATTAACAAACATTAAG 373
 133 rPheTrpLeuPheIleAlaGlyThrAlaMetGlyValAlaSerLeuPhe 150
 374 CTTCTGACTCTTACCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423
 150 LaProGlyLysAspGlyGluLeuGlySerGlyValGlyIleValLeuTyr 166
 424 TG.....GAGCGCGAGCGAGNACGAGTGAACAGCTTAC 458
 167 ProProLeuSerThrArg.....GluAlaGlyTyrSerMetAspLeuAl 181
 459 CCTCCCTTAGCAGGAGACTATCCACCCCTGGAGCTCCGTAGACTTAC 508
 181 aIlePheAlaValHisLeuSerGlyAlaSerIleMetGlyAlaIle 198
 509 CATCTTCTCTTACACCTAGCAGGTCTCTCTCTCTCTCTCTCTCTCTCT 558
 198 smMetIleThrThrPheLeuAsnMetArgAlaProGlyMetThrLeuLeu 214
 559 ATTTCATCCACACATTTATCATATAAAACCCCTGCAATACCAATAC 608
 215 LysValProLeuPheSerTrpSerIlePheIleThrAlaTrpLeuIle 231
 609 CAACGCGCCCTCTCTGCTGATCGCTCAATACACAGCAGCTTACTCT 658
 231 uLeuAlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuTrpAsp 248
 659 CCTATCTCTCCAGCTCTAGCTGCGCATCTACTACTTAACAGAC 708
 248 rAsnProGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 264
 709 GCAACTCTACACACACTTCTTTCGACCCCGGAGGAGGAGACCACTT 758
 265 LeuTyrGluHisIleLeuTrpPhePheGlyHisProGlyValTyrIle 281
 759 CTATACCAACACCTATCTGATTTTTCGTCACCTGAACTTATATCT 808
 281 eIleLeuProGlyPheGlyIleIleSerHisValValSerThrPheSer 297

alignment_block:
US-09-712-768-2 x US-08-998-416-1138

Align seg 1/1 to: US-08-998-416-1138 from: 1 to: 719

```

333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPr 349
4 CTTATATCTAGTACAGCTTATTTTACTCTAGTACATATATATCTTATCTC 53
349 oThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGlyIleSerV 366
104 TAGATTACTAACACCAATATATATATCTATATATATATATATATATAT 153
366 aLGIuPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPhe 382
104 TAGATTACTAACACCAATATATATATCTATATATATATATATATATAT 153
383 ThrValGlyIleValIleThrGlyIleValIleAlaGlnIleAlaAspAr 399
154 ACSTAGTACGTTTAACTGCTAGTACATTAAGCTATATATATATATATAT 203
399 gAlaTyrlHisAspThrTyrTyrValValAlaIleHisPheHisTyrlVal 416
204 AGCATTCATGATACATATATATATATATATATATATATATATATATAT 253
416 eTrleuGlyAlaIlePheAlaIlePheAlaGlyIleTyrlPheTyrlMet 432
254 GTTAGTACGCTGATATCTGATGATGCTGCTGATATATATATATATATAT 303
433 LysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTr 449
304 CTGTGTTTGGTTTAAATATATATATATATATATATATATATATATATAT 353
449 pTrlPhePheIleGlyAlaAsnValIleThrPheProGlnHisPheLeu 466
354 ATTAATTTTCTTAGGCTTATATATATATATATATATATATATATATATAT 403
466 LysGlnGlnIleMetProArgArgTyrIleAspTyrProGlnAlaPheAla 482
404 GTATTATATGCTATACCAAGAACATTCCTGATATATATATATATATATAT 453
483 LeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPh 499
454 GGTTCATATATAGTATCTTATGCTTATATATATATATATATATATATAT 503
499 eLeuPhePheIleValIlePheValTyrThrLeuValAlaGlyArgArg 516
504 AATGTTATATCTTATATATATATATATATATATATATATATATATATAT 553
516 LuThrArgProAsnProTrpGlyIlePhe..... 525
554 ATAAAGTTATATATATATATATATATATATATATATATATATATATATAT 603
526 ..... AlaSprTrlLeuG1 530
604 GAATCAATATATATTTCTTATATATATATATATATATATATATATATAT 653
530 uTrpThrLeuProSerProProAlaHisThrPheGluTrh 544
654 GTTATATATATATATATATATATATATATATATATATATATATATAT 696

```

seq_name: /cogn2_6/protdata/2/lna/6B_COMB.seq:US-08-998-416-187

```

seq_documentation_block:
; Sequence 187, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII

```

```

; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 187:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG10740P
; US-08-998-416-187

```

alignment_scores: Quality: 470.00 Length: 181
Ratio: 3.357 Gaps: 0
Percent Similarity: 77.348 Percent Identity: 48.066

alignment_block: US-09-712-768-2 x US-08-998-416-187 ..

Align seg 1/1 to: US-08-998-416-187 from: 1 to: 663

```

333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleAlaValPr 349
4 CTTGATCTAGAGCTTATTTTACTCTAGTACATATATATTTCTTATCTC 53
349 oThrGlyIleLysIlePheSerTrpIleAlaThrMetTrpGlyIleSerV 366
54 TACTAGATATTAAGTATTTAGTATATATATATATATATATATATATATAT 103
366 aLGIuPheLysSerProMetLeuTrpAlaPheGlyPheMetPheLeuPhe 382
104 TAGATTACTAACACCAATATATATATCTATATATATATATATATATATAT 153
383 ThrValGlyIleValIleThrGlyIleValIleAlaGlnIleAlaAspAr 399
154 ACTTAGTACGTTTAACTGCTAGTACATTAAGCTATATATATATATATAT 203
399 gAlaTyrlHisAspThrTyrTyrValValAlaIleHisPheHisTyrlVal 416
204 AGCATTCATGATACATATATATATATATATATATATATATATATATATAT 253
416 eTrleuGlyAlaIlePheAlaIlePheAlaGlyIleTyrlPheTyrlMet 432
254 GTTAGTACGCTGATATCTGATGATGCTGCTGATATATATATATATATAT 303

```



```

433 LysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTr 449
      |||  ::  |||  ::::::::::::::|||
304 CTGTGTTAGTTAAATTATTAAGAAATTAATCACAATTCATTCG 353
      ::::::::::|||  ::::  |||  ::::::::::|||
449 pThrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeu 466
      ::::::::::|||  ::::  |||  ::::::::::|||
354 ATTAATTTCTTAGCTTAAATATTAATTTCTTCCATGATTCCTAG 403
      ::::::::::|||  ::::  |||  ::::::::::|||
466 lYArgGlnGlyMetProArgArgTyrIleAspTyrProGlnAlaPheAla 482
      ||  ::::|||  ::::|||  ::::|||  ::::|||
404 GTATTAATGATGATCAAGAAAGAAATCCGTATATCCGATCATTCCTA 453
      ||  |||  |||  |||  |||  |||  |||  |||
483 LeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPh 499
      ||  |||  |||  |||  |||  |||  |||  |||
454 GGTGAATTTAGTATCTTCATTTGGTTCCTATTAATCACTATATATCAT 503
      ::::::::::|||  ::::  |||  ::::::::::|||
499 eLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
      ::::::::::|||  ::::  |||  ::::::::::|||
504 AATGTTATTCCTTATATTATTTATGATCAATTAATAATGCT 546

```

seq_name: /cgn2_6/protdata/2/lna/6B_COMB.seq:US-08-998-416-289

seq_documentation_block:

```

; Sequence 289, Application US/08998416
; Patent No. 6239264
;
GENERAL INFORMATION:
;
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPIT
NUMBER OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264Lh Carolina
COUNTRY: USA
ZIP: 27709
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 289:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1241UP
US-08-998-416-289

```

alignment_scores:

Quality: 469.00 Length: 181
Ratio: 3.350 Gaps: 0
Percent Similarity: 77.348 Percent Identity: 48.066

alignment_block:
US-09-712-768-2 x US-08-998-416-289 ...

Align seg 1/1 to: US-08-998-416-289 from: 1 to: 856

```

333 LeuThrGlnGlnSerTyrPheMetLeuAlaThrMetValIleValAla 349
      |||  ::::|||  ::::|||  ::::|||  ::::|||
4 CTGTATCTAGACCTTATTTACTTCACGCTACTATATATTTCTTATTC 53
      ::::::::::|||  ::::  |||  ::::::::::|||
349 oThrGlyIleLysIlePheSerTrpIleAlaTrMetTrpGlyGlySer 366
      |||  ::::|||  ::::|||  ::::|||  ::::|||
54 TACTAGTATTAAGTATTTAGTGAATTAATACTAATTAATGCGTTCAT 103
      ::::::::::|||  ::::  |||  ::::::::::|||
366 aLGlupheLysSerPrometLeuTrpAlaPheGlyPheMetPheLeu 382
      ::::::::::|||  ::::  |||  ::::::::::|||
104 TAAATTAACCAACCAATATTAATCTATTAATCAATTTATTTATTT 153
      ::::::::::|||  ::::  |||  ::::::::::|||
383 ThrValGlyGlyValThrGlyIleValLeuAlaGlnAlaGlyLeuAsp 399
      |||  ::::|||  ::::|||  ::::|||  ::::|||
154 ACTGTAGTGTGTTACTGCTGATGATTTAGCTAATCTATCATTAATG 203
      ::::::::::|||  ::::  |||  ::::::::::|||
399 gAlaTyrHisAspThrTyrTyrValAlaAlaHisPheHisTyrValMet 416
      |||  ::::|||  ::::|||  ::::|||  ::::|||
204 AGCATTCATGATCACTTATTAATTAATGATCACTTATTAATGATTA 253
      ::::::::::|||  ::::  |||  ::::::::::|||
416 eLeuGlyAlaIlePheAlaIlePheAlaGlyIleTyrPheTyrMetPro 432
      |||  ::::|||  ::::|||  ::::|||  ::::|||
254 GTTAGGCTGCTATTCCTATGTTGCTGCTATTAATTAATTAAGTCT 303
      ::::::::::|||  ::::  |||  ::::::::::|||
433 LysPheSerGlyArgAlaPheProGluTrpAlaAlaLysLeuHisPheTr 449
      |||  ::  |||  ::::::::::::::|||
304 CTGTGTTAGTTAAATTATTAAGAAATTAATCACAATTCATTCG 353
      ::::::::::|||  ::::  |||  ::::::::::|||
449 pThrPhePheIleGlyAlaAsnValThrPhePheProGlnHisPheLeu 466
      ::::::::::|||  ::::  |||  ::::::::::|||
354 ATTAATTTCTTAGCTTAAATATTAATTTCTTCCATGATTCCTAG 403
      ::::::::::|||  ::::  |||  ::::::::::|||
466 lYArgGlnGlyMetProArgArgTyrIleAspTyrProGlnAlaPheAla 482
      ||  ::::|||  ::::|||  ::::|||  ::::|||
404 GTATTAATGATGATCAAGAAAGAAATCCGTATATCCGATCATTCCTA 453
      ::::::::::|||  ::::  |||  ::::::::::|||
483 LeuTrpAsnLysValSerSerTyrGlyAlaPheLeuAlaPheAlaSerPh 499
      ||  |||  |||  |||  |||  |||  |||  |||
454 GGTGAATTTAGTATCTTCATTTGGTTCCTATTAATCACTATATATCAT 503
      ::::::::::|||  ::::  |||  ::::::::::|||
499 eLeuPhePheIleValIlePheValTyrThrLeuValAlaGly 513
      ::::::::::|||  ::::  |||  ::::::::::|||
504 AATGTTATTCCTTATATTATTTATGATCAATTAATAATGCT 546

```

seq_name: /cgn2_6/protdata/2/lna/6B_COMB.seq:US-09-385-982-343

seq_documentation_block:

```

; Sequence 343, Application US/09385982
; Patent No. 6262334
;
GENERAL INFORMATION:
;
APPLICANT: ENDECE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
PRODUCTS: II
FILE REFERENCE: CCDA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639

```


NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/518,878B
FILING DATE: 23-AUG-1995
CLASSIFICATION: A43
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-036
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-518-878B-19

alignment_scores:
Quality: 329.50 Length: 91
Ratio: 3.970 Gaps: 1
Percent Similarity: 91.209 Percent Identity: 71.429

alignment_block:
US-09-712-768-2 x US-08-518-878B-19 ..

Align seg 1/1 to: US-08-518-878B-19 from: 1 to: 277

```
222 SerIlePheHleThraIleuLeuAlaLeuProValIleuA 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 TCCGTACTATTACAGCCGCTACTGCTCTATTATCAGTACGAGTGC 53
238 agIyAlaIleThrMetLeuLeuThraSPArGAnPheGlyThrPheP 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 CCGAGCATTAATACTACTACTAGACCGCAATTCCTACACGATCTGTCTGA 103
255 heAsnProAlaGlyGlyAspProIleuLeuTyrgInHisIleuTrp 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 TTGATCCCGCTGGAGGAGGAGCGCAATTCCTACACGATCTGTCTGA 153
272 PhePheGlyHisProGlyValTyrlleIleIleLeuProGlyPheGlyTl 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 TTCTTTGGGACCCAGAGTTTATTATCTTATCTCCAGATTTGGAT 203
288 eIleSerHisValIalSerThrPheSer...LysLysProValPheGlyT 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 TATTTCAATGATGTTACTACTACTCCGAAATAAAGAAACCTTGGCT 253
304 yrLeuProMetValTyrlAlaMet 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ATATAGCAATGATGACCAATA 276
```

seq_name: /cgn2_6/ptodata/2/Ina/5A_COMB.seq:US-08-294-522B-19

seq_documentation_block:
; Sequence 19, Application US/08294522B
; Patent No. 574166

GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: Compositions and Methods for the
TREATMENT OF OBESITY
TITLE OF INVENTION: Treatment of Body Weight Disorders, Including Obesity
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,522B
FILING DATE: 23-AUG-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-015
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-294-522B-19

alignment_scores:
Quality: 329.50 Length: 91
Ratio: 3.970 Gaps: 1
Percent Similarity: 91.209 Percent Identity: 71.429

alignment_block:
US-09-712-768-2 x US-08-294-522B-19 ..

Align seg 1/1 to: US-08-294-522B-19 from: 1 to: 277

```
222 SerIlePheHleThraIleuLeuAlaLeuProValIleuA 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 TCCGTACTATTACAGCCGCTACTGCTCTATTATCAGTACGAGTGC 53
238 agIyAlaIleThrMetLeuLeuThraSPArGAnPheGlyThrPheP 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 CCGAGCATTAATACTACTACTAGACCGCAATTCCTACACGATCTGTCT 103
255 heAsnProAlaGlyGlyAspProIleuLeuTyrgInHisIleuTrp 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 TTGATCCCGCTGGAGGAGGAGCGCAATTCCTACACGATCTGTCTGA 153
272 PhePheGlyHisProGlyValTyrlleIleIleLeuProGlyPheGlyTl 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 TTCTTTGGGACCCAGAGTTTATTATCTTATCTCCAGATTTGGAT 203
288 eIleSerHisValIalSerThrPheSer...LysLysProValPheGlyT 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 TATTTCAATGATGTTACTACTACTCCGAAATAAAGAAACCTTGGCT 253
304 yrLeuProMetValTyrlAlaMet 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ATATAGCAATGATGACCAATA 276
```


OM of: US-09-712-768-2 to: EST:* out-format: pfs

Date: Dec 22, 2001 11:45 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame-p2n.model -DEV=xlp
-O/cg12.1/USPTO.spool/US09712768/runat.21122001.093454.4258/app-query.fasta_1.722
-DE=EST -QPM=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LODPL=0.000 -LODPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELETE=7.000 -START=1 -MATRIX=blissum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USPR=US09712768.@CGNL_1_4287 -NCPU=6
-ICPU=3 -LOGLOG -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

Query: US-09-712-768-2

Query length: 557

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 2217.460000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est2:LI15425	1242.50	1848.30	8.7e-94	1560	LI15425 MARSSTCYA Tammar wallaby
gb_est1:AU175705	1028.50	1530.95	4.1e-76	1054	AU175705 AU175705 Marsupinaeus
gb_est1:AU175493	980.50	1459.35	4.0e-72	1012	AU175493 AU175493 Marsupinaeus
gb_est1:AU167717	946.50	1407.46	3.1e-69	1124	AU167717 OI-br-ad CDN
gb_est1:BE468621	905.50	1347.40	6.9e-66	958	BE468621 IPHdk00496 Head kidney
gb_est1:AU175577	904.50	1344.34	1.0e-65	1146	AU175577 Marsupinaeus
gb_est1:BE574185	890.50	1324.84	1.2e-64	961	BE574185 IPHdk00496 Brain CDNA
gb_est1:BE784814	875.50	1304.19	1.3e-63	784	BE784814 AU395912 Sea urchin
gb_est1:AU395912	875.50	1299.46	3.2e-63	805	AU395912 AU395912 dfzf426 Gallu
gb_est2:BE783431	864.50	1287.28	1.5e-62	822	BE783431 SEAMC003438 Sea urchin
gb_est1:AU183416	864.00	1282.97	2.7e-62	1235	AU183416 AU183416 Cyprinus car
gb_est2:BE784767	863.50	1266.17	1.8e-62	787	BE784767 SEAMC004724 Sea urchin
gb_est2:BI101730	861.50	1261.01	3.4e-62	1007	BI101730 602887509F1 NCI CGAP
gb_est1:AU169837	858.50	1278.43	4.8e-62	809	AU169837 AU169837 OI-br-ad CDNA
gb_est1:BE468572	856.50	1274.39	8.1e-62	911	BE468572 IPHdk00346 Head kidney
gb_est2:BE780373	854.00	1272.55	1.0e-61	798	BE780373 SEAMC000330 Sea urchin
gb_est2:BE6969312	853.00	1269.55	1.5e-61	870	BE6969312 602836784F1 NCI CGAP
gb_est2:BE782703	851.50	1268.09	1.8e-61	795	BE782703 SEAMC002660 Sea urchin
gb_est2:BI147837	849.50	1264.81	2.8e-61	821	BI147837 602912723F1 NCI CGAP
gb_est1:AU393402	849.50	1263.77	3.1e-61	925	AU393402 AU393402 dfzf426 Gallu
gb_est1:AU170132	848.50	1262.72	3.8e-61	930	AU170132 AU170132 OI-br-ad CDNA
gb_est2:BE404674	847.50	1260.78	4.5e-61	902	BE404674 602420835F1 NIH-MGC_94
gb_est2:BI1219070	847.50	1258.98	6.0e-61	824	BI1219070 602938866F1 NCI CGAP
gb_est2:BE781970	845.50	1254.41	1.0e-60	964	BE781970 SEAMC001927 Sea urchin
gb_est2:BI101747	842.50	1252.70	1.3e-60	988	BI101747 602887509F1 NCI CGAP
gb_est2:BI103029	842.50	1252.70	1.3e-60	988	BI103029 602887509F1 NCI CGAP
gb_est2:BI155077	841.00	1251.26	1.6e-60	900	BI155077 602802925F1 NIH-MGC_94
gb_est2:BE6969312	839.00	1248.15	2.3e-60	912	BE6969312 602836784F1 NCI CGAP
gb_est1:AU393803	838.50	1248.48	2.2e-60	810	AU393803 AU393803 dfzf426 Gallu
gb_est2:BE783301	834.50	1242.32	4.9e-60	821	BE783301 SEAMC003258 Sea urchin
gb_est1:AU396852	834.50	1241.86	5.2e-60	866	AU396852 AU396852 dfzf426 Gallu
gb_est2:BE783508	832.50	1239.93	6.7e-60	766	BE783508 SEAMC003465 Sea urchin
gb_est1:AU393370	832.50	1239.51	7.1e-60	804	AU393370 AU393370 dfzf426 Gallu
gb_est2:BE7835260	829.50	1235.22	1.2e-59	785	BE7835260 SEAMC005217 Sea urchin
gb_est2:BE6969264	828.00	1231.48	2.0e-59	953	BE6969264 AU090796 F3F1 NCI CGAP
gb_est1:AU090796	827.00	1229.43	2.6e-59	1080	AU090796 AU090796 F3F1 NCI CGAP
gb_est2:BE783744	824.00	1224.95	3.7e-59	824	BE783744 SEAMC003701 Sea urchin
gb_est2:C23349	823.50	1224.90	4.6e-59	914	C23349 C23349 Japanese flounder
gb_est2:BE783472	821.50	1223.00	5.9e-59	806	BE783472 SEAMC003429 Sea urchin
gb_est2:BE784368	820.50	1222.08	6.6e-59	754	BE784368 SEAMC004325 Sea urchin
gb_est1:BE469917	815.00	1211.84	2.5e-58	948	BE469917 IPHdk01217 Head kidney

gb_est2:BE6964439 + 811.00 1206.31 5.0e-58 898 + BE6964439 602832084F1 NCI CGA
gb_est2:BE6967132 + 811.00 1206.26 5.0e-58 904 + BE6967132 602833803F1 NCI CGA
gb_est2:BE6969596 + 810.50 1206.59 4.8e-58 799 + BE6969596 602836919F1 NCI CGA
gb_est1:AV406294 + 808.50 1203.97 6.7e-58 765 + AV406294 AV406294 Bombyx mor

seq_name: gb_est2:LI15425

seq_documentation_block:

LOCUS LI15425 1560 bp mRNA EST 30-MAR-1995
DEFINITION MARSSTCYA Tammar wallaby mammary gland Macropus eugenii CDNA
similar to transfer RNA-Ser (UCN), mRNA sequence.

ACCESSION LI15425
VERSION LI15425.1 GI:311034

KEYWORDS EST.
SOURCE tammar wallaby.
ORGANISM Macropus eugenii

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.

REFERENCE 1 (bases 1 to 1560)
Collet,C. and Joseph,R.

TITLE The identification of nuclear and mitochondrial genes by sequencing

JOURNAL Randomly chosen clones from a marsupial mammary gland cDNA library
Biochem. Genet. 32, 181-190 (1993)

MEDLINE 95085593
COMMENT Contact: Collet,C. and Joseph,R.

FEATURES
Location/Qualifiers
1..1560

source
1..1560
/organism="Macropus eugenii"
/db_xref="taxon:9315"
/clone_lib="Tammar wallaby mammary gland"
/note="Marsupial mammary gland cDNA library"

BASE COUNT 436 a 392 c 252 g 480 t

ORIGIN

alignment_scores:

Quality: 1242.50 Length: 520
Ratio: 3.146 Gaps: 10
Percent Similarity: 75.962 Percent Identity: 48.846

alignment_block:

US-09-712-768-2 x LI15425 ..
Align seq 1/1 to: LI15425 from: 1 to: 1560

38 AAlaGlyValValGlyPheLeuSerValLeuPheThrValGlyMetArgLeu 54

13 GCAGGATATGAGGA.....ACTGCCCTAAGCTTGCTCATTCGTGC 53

54 UGULeuMetAspProGlyValGlnTyrMetCysLeuGluGlyAlaArgL 71

54 AGAAGCTTGCCCAACCGCG..... 72

71 euilleaAspAlaSerGlnThrCysThrAlaAsnGlyHisLeuTyrPsn 87

73ACCTGATCGGACGCGATTAATGATATGATATGATGCG 102

88 ValMetValThrTyrHisGlyLeuMetMetPhePheValGlyLeu 104

103 GTATTTGTACCGCGCGCATGCTTGTATGATATGATATGATGCG 152

104 oAlaLeuPheGlyGlyPheGlyValGlyValGlyValGlyValGlyVal 121

153 TATCATATTTGAGAGATTTGGCAACGATTTGGTACCTTATATATGCTG 202

121 lapRoAspMetAlaPheProArgMetAsnAsnLeuSerPheTyrPhe 137

203 CACCGCATATGACATTCGCCAGGATTAATTAACATTAACCTTGACTTTTA 252

138 lIleAlaGlyTyrAlaMetGlyValAlaSerLeuPheAlaProGlyGlyAs 154

253 CCCCACCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 287


```

57 etasproglyvalglntrymetcysleuigllyalalaleuileala 73
   ::::::::::::::::::::
103 GTCACACGAGA.....AGACTTATGGA 125
74 Aspalaserglnthrcysthralalasnglyhisteutrpasvalmetva 90
   ::::::::::::::::::::
126 GAT.....GACCAATCTATATGATGAGT 151
90 lthtryhsglylleulemetphepvalglylleproalaleu 107
   ::::::::::::::::::::
152 TACAGCTCATGCACTTTGTTATATTTCTTATAGTATACCTATATTA 201
107 heglyglypheglyasntryleuemetproleuglnileglyalaproasp 123
   ::::::::::::::::::::
202 TTGGAGATTGGAACTGGTATGCTTAAATATATAGCAGCAGCCTGAT 251
124 Metlaphheproargmetasnasleuserpheprhephelealeagl 140
   ::::::::::::::::::::
252 ATGGCAATCCACGAAATGAATATATAGCTTTGGCTGCTGCTCCTTC 301
140 ythrAlameglyvalAlaserleuphealaproglylyaspglylnl 157
   ::::::::::::::::::::
302 T.....CTTACTCTCTCTTATCCAGAGCAATAGTTGAA 336
157 euglyserglyvalglytrpvalleutyrproproleuserthrArg... 172
   ::::::::::::::::::::
337 GAGAGATTGAGAACGATGAGCTGTACCTCCTTATACGCTAGATC 386
173 ...Gluhaglytyrsermetasplealalephealavalhisteuse 188
   ::::::::::::::::::::
387 GCACATCCAGGGGCTGTGATGATTAGAAATTTTTCATCTACCTTTAGC 436
188 rglAlaserSerllemetglyalaleasmetllethrThPheleua 205
   ::::::::::::::::::::
437 AGGGGTTTCATCATTTTAGGGGCGCTAAATTTATACAACTAGTATTA 486
205 smetargalaproglymetthrleuhslyvalproleupeserTIP 221
   ::::::::::::::::::::
487 ATATACATCTACGATGATGATGACCGCAATACCGTATGCTGCTG 536
222 SerllepethrAlatrpleuileleulealaleuprovalleual 238
   ::::::::::::::::::::
537 GCAGCTTTTACTGCACTTTTACTCTTATCTTACTTACTGTTCTAGC 586
238 aglyAlalethrmetleuethrAspArgasnPheglyThrThPhep 255
   ::::::::::::::::::::
587 AGAGGATATACATCTTCTTACGATCGAATCTTAATACCTCATTTT 636
255 heasnproalaglyglyaspProilleuetyrglnhisteutrip 271
   ::::::::::::::::::::
637 TTGACCCAGCGGGTGGAGAGACCCAGCTTATCAACATTTATTTCTG 686
272 PhepheglyhispProgluvaltyrlelleleleupProglyPheglyl 288
   ::::::::::::::::::::
687 TTCTTCGGGACCCAGAGCTTATATTTCTTCTCCCGCTTGAAT 736
288 eliserhisvalalserThrPheSer...LysLysprovalPheglyl 304
   ::::::::::::::::::::
737 AATCTCACATCATCATGTCAGAGCTCGGGTAAAAAAGAGCTTTGCTA 786
304 ytleuprometvaltyrAlamevalalelleglyvalleuglyPheval 320
   ::::::::::::::::::::
787 CTTTAGGATTAATTTATGCTATATTTGGCATTTGAGTTTATGATTT 836
321 ValtrpAlahshismetlyrThrValglymetSerleuthrglninse 337
   ::::::::::::::::::::
837 GTGGGGGCTCACATATATTTACAGTAGAATGATGTGATACACAGC 886
337 rtyrPheMetleualatThrMetVallealavalProThrGlylleuysI 354
   ::::::::::::::::::::
887 TTACTTTTACATCAGCAACAATATTTATTTGCTGCTCCTACGGAATTAAT 936
354 lepheserTIPleAlaThrMetTIP 362

```

```

937 CTTTACTTGCCTTGGAGACCTCCATCG 962
seq_name: gb_est1:AUI67717
seq_documentation_block: 1124 bp mRNA EST 29-JAN-2001
LOCUS AUI67717 1124 bp mRNA EST 29-JAN-2001
DEFINITION AUI67717 O1-br-ad cDNA oryzias latipes cDNA clone br0955, mRNA
sequence.
ACCESSION AUI67717
VERSION AUI67717.1 GI:12589786
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 1124)
AUTHORS Mita,K., Ishikawa,Y. and Yamauchi,M.
TITLE Establishment of cDNA database of medaka, Oryzias latipes
JOURNAL Unpublished (2001)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmits@nirs.go.jp
method:uni-directional sequence direction:sequenced from T3 primer
(5' -> 3').
FEATURES
source
Location/Qualifiers
1..1124
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="br0955"
/clone_1bp="O1-br-ad cDNA"
/sex="female/male mixed"
/rissu_type="brain"
/dev_stage="adult"
BASE COUNT 292 a 284 c 210 g 338 t
ORIGIN
alignment_scores:
Quality: 946.50 Length: 344
Ratio: 3.392 Gaps: 4
Percent Similarity: 81.105 Percent Identity: 53.779
alignment_block:
US-09-712-768-2 x AUI67717 ..
Align seg 1/1 to: AUI67717 from: 1 to: 1124
207 ArgAlaproglymetthrleuhslyvalproleupeserTIPserl 223
   ::::::::::::::::::::
2 AAACCCCGCCATCTCTCAATACCAACCCCTTATATTTGTTGAGCGGT 51
223 epheliehrAlatrpleuileleulealaleuprovalleualagly 240
   ::::::::::::::::::::
52 ACTAATACACGCGAGATATTTACTTCTCTCTCCCTGCTGAGCTGAG 101
240 lalThrmetleuethrAspArgasnPheglyThrThPhepheasn 256
   ::::::::::::::::::::
102 GGATCACTTACTCTTACAGACCGCAACCTTAATACCAACCTTCTTCGC 151
257 proAlaglyglyaspProilleuetyrglnhisteutriphep 273
   ::::::::::::::::::::
152 CCGAGAGAGAGGAGCCCTATTTCTACCAACACCTTTTCTGATTTCT 201
273 eglyhispProgluvaltyrlelleleupProglyPheglylleu 290
   ::::::::::::::::::::
202 TGCGCACCTGAAGTTTATATTTCTATTTCTCCCGGCTTGGAGATTT 251

```


seq_name: gb_estl:BE574185

seq_documentation_block:

LOCUS BE574185 mRNA EST 15-AUG-2000
DEFINITION Iprmo101227 Brain cDNA library Ictalurus punctatus cdna 5' similar to Mitochondrion tRNA(Ser) [UCN], tRNA(asp) and cytochrome c oxidase subunits I and II genes, mRNA sequence.
ACCESSION BE574185
VERSION BE574185.1 GI:9824230
SOURCE channel catfish.
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.
1 (bases 1 to 961)
Ju,Z., Kasai,A., Kocabas,A., Patterson,A., Li,P., Cao,D., Dunham,R. and Liu,Z.
Transcriptome analysis of channel catfish: I. genes and expression profiles from the brain
Unpublished (2000)
CONTACT Contact: Liu ZJ
The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel.: 334 844 4054

FEATURES
 Source .
 Location/Qualifiers
 1..961
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Brain cDNA library"
 /note="Organ: Brain; Vector: pSPORT1; Site_1: NotI
 Site_2: SalI"
 BASE COUNT 247 a 266 c 166 g 282 t
 ORIGIN
 alignment_scores:
 Quality: 890.50 Length: 301
 Ratio: 3.425 Gaps: 1
 Percent Similarity: 86.379 Percent Identity: 55.814
 alignment_block:
 US-09-712-768-2 x BE574185 ..
 Align seg 1/1 to: BE574185 from: 1 to: 961
 207 ArgAlaProGlyMetThrLeuHisLysValProLeuPheSerTrpSerI 223
 :::
 1 AAGCCCCCGCAATCTCAAAATCAAAACCCCTTATTTGTCAGCCGT 50
 223 ephetherlartpleuileleuAlalaLeuProValleuAlaIya 240
 :::
 51 CCTATTACAGCCGTCCTCTACTACTACTCTCCACTTTTACCCCTG 100
 240 laileThrMetLeuLeuThrAspArgAsnPhelGlyThrThrPheAsn 256
 :::
 101 GATACATATCTTCTTAAACAGACCGAAACTTAATCTACTCTTTGAC 150
 257 ProAlaGlyGlyAspProIleLeuTyroHisIleLeuTrpPhePh 273
 :::
 151 CCGCAGAGAGG.GGGGACCCCATCTTTTACCAACCTTTTCGATCTT 199
 273 eGlyHisProGluValTyroIleIleIleLeuProGlyPheGlyIle 290
 :::
 200 TGGGACCCAGAGATATCATTTAAATTTTCCAGGCTTGGCATATCT 249
 290 erHisValValSerThrPheSer..LysLysProValPheGlyTyro 305
 :::
 250 CCCAATTTGTTGCCCTACTACGACGCAAAAAGAACCATCTGCGTAT 299
 306 ProMetValTyroAlaMetValAlaIleGlyValLeuGlyPheVal 322
 :::
 300 GGAATAGCTGAGCTATATATAGCATTTGGCTTTAGGCTTATGTATG 349
 322 AlaHisHisMetTyroThrValGlyMetSerLeuThrGlnGlnSer 359
 :::
 350 AGCATCATCATATTTCACTAGAGGAAAGACGTGACATCGACATATT 399
 359 hemetLeuAlaThrMetValIleAlaValProThrGlyIleLysIle 365
 :::
 400 TTACTATCTGCACATATATTTATTCATATCCCACTGGGGTCAAAAGT 449
 365 SerTrpIleAlaThrMetTrpGlyGlySerValGluPheLysSerPro 372
 :::
 450 AGCTGACTTGGCTACCCCTACAGGCGGCATATTAATGAGAAACTCC 499
 372 LeuTrpAlaPheGlyPheMetPheLeuPheThrValGlyValThrG 389
 :::
 500 GCTCGACCCCTAGGTTTCATCTTCCTTACGTTAGTGGGGGCTTAAC 549
 389 LysIleValLeuAlaGlnAlaGlyLeuAspArgAlaTyroLysAsp 405
 :::
 550 GCATTGTACTAGCCACATCATCTCTACACATTTGTTCTACATGACAC 559
 406 TyrValValAlaHisPheHisTyroValMetSerLeuGlyAlaIle 422

```

|||||
600 TATGTATACCCCATTTCCATTAATGTATCAATGAGGCGTATTTGC 649
422 ATLePheAlaGlyLeuTyrPheYrMetProLysPheSerGlyArgAlaP 439
650 TATTATGGGGGCGCTTCCTGCTACGTGATTCCTCTTACAGCATACCA 699
439 heProGluTrrPAlaAlaLysLeuHisPheTrrPrrPhePheAlaGlyAla 455
700 TACGACGACACTGACAAATAATTCACCTTGGAGACTATATTGGGGGCTA 749
456 AsnValThrPhePheProGlnHisPheLeuGlyArgGlnGlyMetProAr 472
750 AACTCAGCTTCTTCCCAACACTTCTAGCCCTCGCTGCATACACG 799
472 gArGTyrLlAsPrrPrrGlnAlaPheAlaLeuTrrPrrAlaLysValSer 489
800 ACGATCTACAGACTACCCGATGCTACTCATATGAAACATGATCTCTT 849
489 eTrrGlyAlaPheLeuAlaPheAlaSerPheLeuPheAlaValLe 505
850 CAATGGGCTCCTGCTGCTCCCTAGTACAGTGTATATCTGATATTT 899
506 Phe 506
900 TAT 902

seq_name: gb_est2:BG784814

seq_documentation_block:
LOCUS BG784814 784 bp mRNA EST 20-MAY-2001
DEFINITION SEAMC004771 Sea urchin primary mesenchyme cell cDNA library
Strongylocentrotus purpuratus cDNA clone PC_0023_A1_H09_MR 5', mRNA
sequence.
ACCESSION BG784814 GI:14155827
VERSION BG784814.1 GI:14155827
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinodermata; Echinodermata; Echinodermata;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 784)
Zhu,X., Mahaltras,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
Development 128 (2001) In press
JOURNAL Contact: Ettensohn CA
COMMENT Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.

FEATURES
source
1..784
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone_lib="PC_0023_A1_H09_MR"
/clone_lib="Sea urchin primary mesenchyme cell cDNA
library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
dt priming from poly A+ RNA, directionally cloned"

BASE COUNT 184 a 184 c 157 g 256 t 3 others

alignment_scores:
Quality: 875.50 Length: 260
Ratio: 3.908 Gaps: 1

```

```

Percent similarity: 86.154 Percent identity: 61.154
alignment_block:
US-09-712-768-2 x BG784814
Align seg 1/1 to: BG784814 from: 1 to: 784

234 LeuProValLeuAlaGlyAlaLeuPheLeuLeuThrAspArgAsn 250
5 ATTTCCGGGTGACCCACCGCGTCCGCAATGCTTTCACAGATCGTAAT 54
250 eGlyThrThrPhePheAsnProAlaGlyGlyGlyAspProLeuTyrG 267
55 AACACAACTTCTTGAACCTGACGAGAGGGAGATCAATTCGTTTC 104
267 LHisLLeuThrPhePheGlyHisProGluValTyrLLeuLeuLeu 283
105 AACACCTATTCGTTTGTGGACACCCCGAGGTGATATTCCTTATTC 154
284 ProGlyPheGlyLLeuSerHisValValSerThrPheSer...LysLy 299
155 CCGGATTTGGTATGATCTACACGTTTATAGCTCACTGCTGTAACG 204
205 AGAGCCTTCGGATACCTGGGATGTTTATGCATGATTCGATAGAG 254
299 sProValPheGlyTyrLeuProMetValTyrAlaMetValAlaLeu 316
316 AlLeuGlyPheValValTrrPAlaHisHisMetTrrThValGlyMetSer 332
255 TTTTGGATTCCTTCTGCGGCCACCAATGTTTACAGTGAATGAT 304
333 LeuThGlnGlnSerTyrPheMetLeuAlaThrMetValLeuAlaValPr 349
305 GTTGATACACGACACTACTCTACCTCGCCACAGATGATATTGCGTCC 354
349 ThrGlyLLeuLysLLeuPheSerTrpLeuAlaThrMetTrpGlyGlySer 366
355 AACAGGATTAAGGTTTTCAGATGAATGACCAAGCTCCAGAGTCAATC 404
366 AlGluPheLysSerProMetLeuTrrPAlaPheGlyPheMetPheLeu 382
405 TACAAAGAAACTCCCTATATATAGACCTTGGAATTTGATTTTATTT 454
383 ThrValGlyGlyValThrGlyLLeuValLeuAlaGlnAlaGlyLeuAspAr 399
455 ACATTAGAGAGACTCACAGGTATGTTGCTTGGCAATTCCTCCATGAG 504
399 gAlaTyrHisAspThrTrrTrrValValAlaHisPheHisTrrValMet 416
505 TGTTCTCATGATACCTACTAGCTGAGTCACTTCACTAGTCACTGCT 554
416 eLeuGlyAlaLeuPheAlaLeuPheAlaGlyLeuTyrPheTyrMetPro 432
555 CATGGGGGCTGATTTTGCATCTTCGCTGGTTCACACAGGTTTCC 604
433 LysPheSerGlyArgAlaLeuProGluTrrPAlaAlaLysLeuHisPheTr 449
605 CTTCTGCGGTTATAGCTACCAATATAGAGAAAGGTCCACTGCTT 654
449 rThrPhePheLeuLeuGlyAlaAsnValThrPhePheProGlnHisPheLeu 466
655 CATTAAGTTTGTGGAGTCACCTTAACCTTTTTCCTCAACCTTTCTTA 704
466 LArgGlnGlyMetProArgArgTrrLLeuAspTrrPrrGlnAlaPheAla 482
705 GCTAGCCCGGNAATGCCACGACNGTACTCAGACTATCCAGAGCCCTATACA 754
483 LeuTrrPrrAlaLysValSerSerTrrGlyAla 492
755 CTTTGAATATCATCTNCTCAATTTGATACA 784

seq_name: gb_est1:AJ395912

```

seq_documentation_block:	LOCUS	AJ395912	805 bp	mRNA	EST	25-JAN-2001
DEFINITION	AJ395912	dkfz426	Gallus gallus	cDNA clone 24j17r1	mRNA sequence.	
ACCESSION	AJ395912					
VERSION	AJ395912.1	GI:7127314				
KEYWORDS	EST.					
SOURCE	chicken.					
ORGANISM	Gallus gallus					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.					
REFERENCE	1 (bases 1 to 805)					
AUTHORS	Abdrakhmanov,I., Lodgjin,D., Gerolt,P., Arakawa,H., Law,A., Plachy,J., Korn,B. and Buerstedde,J.M.					
TITLE	A large database of chicken bursa ESTs as a resource for the analysis of vertebrate gene function					
JOURNAL	Genome Res. 10 (12), 2062-2069 (2000)					
MEDLINE	20568495					
COMMENT	Contact: Buerstedde JM Cellular Immunology Heinrich-Pette-Institute Martinistr. 52, 20251 Hamburg, Germany Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html . Location/Qualifiers 1. 805					
FEATURES	source					
	/organism="Gallus gallus" /strain="CB" /db_xref="taxon:9031" /clone="24j17r1" /clone_lib="dkfz426" /tissue_type="Bursa of Fabricius"					
BASE COUNT	214 a	252 c	136 g	203 t		
ORIGIN						
alignment_scores:	Quality: 872.50	Length: 266				
	Ratio: 3.729	Gaps: 2				
Percent Similarity:	87.970	Percent Identity:	61.654			
alignment_block:	US-09-712-768-2 x AJ395912 ..					
Align seg 1/1	to: AJ395912	from: 1	to: 805			
156	GlnLeuGlySerGlyValGlyTrpValLeuTrpProProLeu.....Se 170					
7	GAGCTGGGGCGCGCACAGATGAGCAGTTTACCCCTTTAGCCGGCAA 56					
170	rThrArgGlnAglGlyTrpSerMetAspLeuValAlaIlePheAlaValHisL 187					
57	CuTACCACCGAGCTGGCGGCATCAGTAGACCTAGCCATCTTTTCATTACACT 106					
187	euSerGlyValAspSerIleMetGlyAlaIleAsnMetIleTrpThrPhe 203					
107	TGACAGGGGTTTCCTCCATCTCAGAGCACAACCTTACACACATC 156					
204	LeuAsnMetArgAlaProGlyMetThrLeuHisLysValProLeuPheSe 220					
157	ATCAACATTAACACCCCGCACCTGTCAATACCAACACCCCTATTTCGT 206					
220	rTrpSerIlePheIleThrAlaTrpIleuIleLeuAlaLeuProVal 237					
207	ATGATTCGCGTCTCATTTACTGCGCATCTACTCTCTCTCTTACCGCTCC 256					
237	euaIleGlyAlaIleThrMetLeuLeuThrAspArgAsnPheGlyThrThr 253					
257	TGACAGCTGGGATTACATTAATCTTACCGACCCCAACCTTAACACACA 306					
254	PhePheAsnProAlaGlyGlyValAspProIleLeuTrpGlnHisIleLe 270					
307	TTCTTCGACCCGCTGGAGGAGGAGACCCCAATCTTATACACACACTATT 356					

[illegible]

```

seq_name: gb_est2:BG783541
seq_documentation_block:
LOCUS      BG783541      822 bp      mRNA      1
DEFINITION      SEANUM003498 Sea urchin primary mesenchyme cell cDNA library
Strongylocentrotus purpuratus cDNA clone PC_0018_B1_E11_MR 5', mRNA
sequence.
BG783541
BG783541.1      GI:14154554
EST.
Strongylocentrotus purpuratus.
Strongylocentrotus purpuratus
Eukaryote; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 822)
Zhu,X., Mahaltras,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
Development 128 (2001) In press
Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
Location/Qualifiers
1..822
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="PC_0018_B1_E11_MR"
/clone_id="Sea urchin primary mesenchyme cell cDNA
library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"

```

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
at priming from poly A+ RNA, directionally cloned"

BASE COUNT 189 a 203 c 166 g 263 t 1 others

alignment_scores:

Quality: 864.50 Length: 269
Ratio: 3.726 Gaps: 3
Percent Similarity: 86.245 Percent Identity: 61.338

alignment_block:
US-09-712-768-2 x BG783541 ..

Align seg 1/1 to: BG783541 from: 1 to: 822

```

151 ProGlyGlyAspGlyGlnLeuGlySerGly.....ValGlyTr 163
    |||:||||| ||| ||| |||
4 CCGCAGAGTACCGGTCGGAATTCGCCGACGCCGCGCGCTG 53
    |||:||||| |||:||||| |||:||||| |||:
163 pValLeuTyProProLeuSerThrArg.....GluAlaGlyTyrSer 178
    |||:||||| |||:||||| |||:||||| |||:
54 AACATATACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 103
    |||:||||| |||:||||| |||:||||| |||:
178 etAspLeuAlaIlePheAlaValAlaHisLeuSerGlyAlaSer 194
    |||:||||| |||:||||| |||:||||| |||:
104 TTGATTAGCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 153
    |||:||||| |||:||||| |||:||||| |||:
195 GlyAlaIleAsnMetIleThrThrPheLeuAsnMetArgAlaPro 211
    |||:||||| |||:||||| |||:||||| |||:
154 GCCCAATTAATTAATTAACACAAATTAATTAATTAATTAATTA 203
    |||:||||| |||:||||| |||:||||| |||:
211 tThrLeuHisLysValProLeuPheSerThrPheIleThrAla 228
    |||:||||| |||:||||| |||:||||| |||:
204 GTCTTTGATCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 253
    |||:||||| |||:||||| |||:||||| |||:
228 rPheUlleLeuLeuAlaLeuProValLeuAlaGlyAlaIleThr 244
    |||:||||| |||:||||| |||:||||| |||:
254 TCTTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 303
    |||:||||| |||:||||| |||:||||| |||:
245 LeuThrAspArgAsnPheGlyThrThrPhePheAsnProAlaGly 261
    |||:||||| |||:||||| |||:||||| |||:
304 CTCACATATGTAATAATTAACACAACTTCTTGACCTCTCAGAG 353
    |||:||||| |||:||||| |||:||||| |||:
261 yAspProIleLeuTyGlnHisIleLeuTrpPhePheGlyHisPro 278
    |||:||||| |||:||||| |||:||||| |||:
354 AGATCCAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 403
    |||:||||| |||:||||| |||:||||| |||:
278 aTyTrIleIleIleLeuProGlyPheGlyIleIleSerHisVal 294
    |||:||||| |||:||||| |||:||||| |||:
404 TGTATATTTCTTATCTTACCGGATTTGTATGATCTCACAGT 453
    |||:||||| |||:||||| |||:||||| |||:
295 ThrPheSer...LysLysProValPheGlyTyTrLeuProMet 310
    |||:||||| |||:||||| |||:||||| |||:
454 CACTACTCTGTGTAAGCAGAGCCTTTCTGGAACTGGGGGTGTT 503
    |||:||||| |||:||||| |||:||||| |||:
310 aMetValAlaIleGlyValLeuGlyPheValAlaTrpAlaHis 327
    |||:||||| |||:||||| |||:||||| |||:
504 CATGATTCGATAGAGAGATTAGATTCCTGTCTGGCCACCATAT 553
    |||:||||| |||:||||| |||:||||| |||:
327 yTrThrValGlyMetSerLeuThrGlnGlnSerTyrrPheMet 343
    |||:||||| |||:||||| |||:||||| |||:
554 TTACAGTAGGAAATGATGTATACACACATCTTCACTGCCGAC 603
    |||:||||| |||:||||| |||:||||| |||:
344 MetValIleAlaValProThrGlyIleLysIlePheSerTrp 360
    |||:||||| |||:||||| |||:||||| |||:
604 ATGATTAATTCCTCTCTCAACAGAGATTAANGTTTTCAGATGA 653
    |||:||||| |||:||||| |||:||||| |||:
360 rMetTrGlyGlySerValGlnPheLysSerProMetLeuTrpAla 377
    |||:||||| |||:||||| |||:||||| |||:
654 GCTCCAGAGGTCTAATCTACATGAGAAACCTCTTATTTAGCC 703
    |||:||||| |||:||||| |||:||||| |||:
377 LysPheMetPheLeuPheThrValGlyGlyValThrGlyIleVal 393
    |||:||||| |||:||||| |||:||||| |||:

```

```

|||||:||||| |||:||||| |||:||||| |||:
704 GATTGTATTTTATTTATACATAGGAGACCTCACAGTATGCTTGC 753
    |||:||||| |||:||||| |||:||||| |||:
394 GluAlaGlyLeuAspArgAlaTyrrHisAspThrTyrrValAla 410
    |||:||||| |||:||||| |||:||||| |||:
754 AATTCCTCATTTGACGTTGTCTTCATGATACCTACTACGTGTAG 803
    |||:||||| |||:||||| |||:||||| |||:
410 sPheHis 412
    |||:||||| |||:||||| |||:||||| |||:
804 CTTTCAC 810

```

seq_name: gb_est1:AU183416

seq_documentation_block:

LOCUS AU183416 1235 bp mRNA EST 01-MAY-2001

DEFINITION AU183416 Cyprinus carpio head kidney stimulated by 1ipo-polysaccharide and concanavalin-A Cyprinus carpio cDNA clone

ACCESSION AU183416 GI:13919157

VERSION AU183416

KEYWORDS EST.

SOURCE common carp.

ORGANISM Cyprinus carpio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.

REFERENCE 1 (bases 1 to 1235)

AUTHORS Sakai, M. and Sazan, R.

TITLE Analysis of expressed sequence tags from a Common carp, Cyprinus

carpio, head kidney cDNA library

JOURNAL Unpublished (2001)

COMMENT Contact: Masahiro Sakai

Faculty of Agriculture

Miyazaki University

1-1 nishi gakukenkibanadai, Miyazaki, Miyazaki 889-2192, Japan

Email: m.sakai@cc.miyazaki-u.ac.jp.

Location/Qualifiers

1. 1235

/organism="Cyprinus carpio"

/db_xref="taxon:7962"

/clone="H74"

/clone_1id="Cyprinus carpio head kidney stimulated by

1ipo-polysaccharide and concanavalin-A"

/tissue_type="head kidney stimulated by

1ipo-polysaccharide and concanavalin-A"

BASE COUNT 369 a 301 c 213 g 352 t

alignment_scores:

Quality: 864.00 Length: 306
Ratio: 3.585 Gaps: 5
Percent Similarity: 76.758 Percent Identity: 55.229

alignment_block:
US-09-712-768-2 x AU183416 ..

Align seg 1/1 to: AU183416 from: 1 to: 1235

```

249 AsnPheGlyThrThrPhePheAsnProAlaGlyGlyAspProIle 265
    |||:||||| |||:||||| |||:||||| |||:
230 AACCTTTATACCAATCTTTGACCCGCGAGGAGAGAGACCAATCT 279
    |||:||||| |||:||||| |||:||||| |||:
265 uTyrrGlnHisIleLeuTrpPhePheGlyHisProGlyValTyrr 282
    |||:||||| |||:||||| |||:||||| |||:
280 TTATCAACACTTATTCGATTCCTCGGCCACGAGAGATTATATTC 329
    |||:||||| |||:||||| |||:||||| |||:
282 LeuLeuProGlyPheGlyIleIleSerHisValAlaSerThrPhe 327
    |||:||||| |||:||||| |||:||||| |||:
330 TCTCTCCAGAGATTGCGTATCATCTCCCAATGTTGATGCTATTTAC 379
    |||:||||| |||:||||| |||:||||| |||:
298 LysLysProValPheGlyTyrrLeuProMetValTyrrAlaMetVal 314
    |||:||||| |||:||||| |||:||||| |||:

```

380 AAAAAAGACCGTTCGTTACATAGCAATGCTGAGCCATGATGCCAT 429
 314 eGlyValLeuGlyPheValIValTPraIahisMetYrThraIglYM 331
 430 CGGTCTACAGGTTATTTATGTGTGAGCTCACCACATGTTCACTGTTGGA 479
 331 eSerLeuThrGlnGlnSerYrPheMetLeuAlaThrMetValIleAla 347
 480 TAGAGTAGACACACGAGCATATTTACATCCGCAACAATATATTCGA 529
 348 ValProThrGlyIleYsIlePheSerTrpIleAlaThrMetTrpGlyG 364
 530 ATTCACACAGGTTGTAAGATATTAGCGATAGCCACACTCCACGAGG 579
 364 ySerValGluPheYsSerProMetLeuTrpAlaPheGlyPheMetPhe 381
 580 ATCCATTAAATGAGAAACACCAATCTATGAGCTTAGGGTTATTTTCC 629
 381 eupPheThrValIglYValIThrGlyIleValIleuAlaGlnAlaIleu 397
 630 TATTCACAGTAGGGGAGCTCAGAGGATGCTTCATCCAAATCATCACCT 679
 398 AspArgAlaIYrHisAspThrYrYrValValAlaIahisPheHisYrVa 414
 680 GATATTGCTCTCAGACACTTACTAGTAGAGACACTTCCACTATGT 729
 414 lMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleYrPheYrYM 431
 730 ACTATCATATGGTGTGTATTTGCTATATAGGACCTTTGGCACTTGAT 779
 431 eTrProYsPheSerGlyArgAlaPheProGluTrpAlaAlaYsLeuHis 447
 780 TCCCCCTTAACTGATATCTCTCCACAGCAGCCTGAACAAATATCCAC 829
 448 PheTrpPhePhePheIleGlyAlaIasValIThrPhePheProGlnHisPh 464
 830 TTTCGAGTTATGTTATGAGTTAACTCAGCTTCTTCCCAACACCTT 879
 464 eLeuGlyArgGlnGlyMetProArgArgYrIleAspYrProGluAlaP 481
 880 CCTAGGCTTAGCCGGTATGCCAGCGCATCTCCGACATATCCAGATGCT 929
 481 hEaIleuTrpAsnYsValSerSerYrGlyAlaIlePheAlaIlePheAla 497
 930 ACGCCCTGGAATATAGATATCATCATGATGATCGCTAATTTCC..... 973
 498 SerPheLeuPhePheIleValIlePheValYrThrLeu.....Va 511
 974CTAGTAGCAGTAAATCATATCTATTTATCTATGAGAGCCCTT 1017
 511 lAlaGlyArgArgIuThrArgProAsnProTrpGlyIuPheAlaAspT 528
 1018 CGCCGCTTAAGACAGAACTA.....T 1040
 528 hLeuGluTrpThrLeuProSer.....ProPro 537
 1041 CTGTAGACTAAGACCAACAATGTAGATGACTTACAGGCTGCCCTCT 1090
 538 ProAlaHisThrPheGlu 543
 1091 CCTTATCACACATACGAG 1108
 seq_name: gb_est2:BG784767
 seq_documentation_block: 787 bp mRNA EST 20-MAY-2001
 LOCUS BG784767
 DEFINITION SPAMUC004724 Sea urchin primary mesenchyme cell cDNA library,
 Strongylocentrotus purpuratus cDNA clone PC_0023_Al_D09_MR 5', mRNA
 sequence.
 ACCESSION BG784767
 VERSION BG784767.1 GI:14155780
 KEYWORDS EST.
 SOURCE Strongylocentrotus purpuratus.

ORGANISM Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoidea; Echinoidae; Echinacea; Echinoida;
 Strongylocentrotidae; Strongylocentrotus.
 REFERENCE 1 (bases 1 to 787)
 AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
 Etlensohn,C.A.
 TITLE A large scale analysis of mRNAs expressed by primary mesenchyme
 cells of the sea urchin embryo
 JOURNAL Development 128 (2001) In press
 COMMENT Contact: Etlensohn CA
 Dept. Biol. Sci.
 Carnegie Mellon University
 4400 Fifth Avenue, Pittsburgh, PA 15213, USA
 Tel: +1 412 268 5849
 Email: etlensohn@andrew.cmu.edu.
 FEATURES
 source
 1..787
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone="PC_0023_Al_D09_MR"
 /clone_lib="Sea urchin primary mesenchyme cell cDNA
 library"
 /tissue_type="embryo"
 /cell_type="primary mesenchyme cells"
 /lab_host="E.coli"
 /note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
 dt priming from poly A+ RNA, directionally cloned"
 BASE COUNT 184 a 186 c 162 g 253 t 2 others
 ORIGIN
 alignment_scores
 Quality: 863.50 Length: 248
 Ratio: 3.998 Gaps: 1
 Percent Similarity: 87.097 Percent Identity: 62.500
 alignment_block:
 US-09-712-768-2 x BG784767 ..
 Align seg 1/1 to: BG784767 from: 1 to: 787
 246 ThrAspArgAsnPheGlyTrpThrPheAsnProAlaGlyGlyAs 262
 41 ACAATGCTAAATAAACACACACTTTCTTGACCTCCAGAGGGGAGCA 90
 262 pProIleuTrpGlnHisIleuTrpPhePheGlnHisProGluVal 279
 91 TCCAAATCTGTTTCACACCCATCTGTGTTTGGACACCCGAGGTGT 140
 279 YrIleIleIleuProGlyPheGlyIleIleSerHisValIleSerThr 295
 141 ATATTCTTAATCTACCGGATTTGGTATGATCTCAGCAGTTATAGCTC 190
 296 PheSer...LysLysProValPheGlyTrpLeuProMetValYrAlaMe 311
 191 TACTCTGTATAGCAGAGCCCTTGATACCTGGGATGGTTATGCGAT 240
 311 tValAlaIleGlyValLeuGlyPheValIValTPraIahisMetYrT 328
 241 GATTGCAATAGAGTTTATGATCTTGTCTGTGGCCACCAATATGTTTA 290
 328 hValGlyMetSerLeuThrGlnGlnSerYrPheMetLeuAlaThrMet 344
 291 CAGTAGGAATGATGTTGATACAGACGATATCTTACCTGCCGACGATG 340
 345 ValIleAlaValProThrGlyIleYsIlePheSerTrpIleAlaThrMe 361
 341 ATATTGCTGTTCCACACAGGATTAAGTTTTCAGATGAAATGCAACGCT 390
 361 tTrpGlyGlySerValGluPheYsSerProMetLeuTrpAlaPheGlyP 378
 391 CCAAGGGCTTAATCTACAAATGAGAACTCCCTTAATATGAGCTTGAGAT 440

```

378 hemetPheLeuPheThrValGlyValThrGlyIleValLeuAlaGln 394
|||||
441 TTGATATTTTATTACATTTAGAGAGACTCACAGATATGTTCTGCCAAT 490
|||||
395 AlAGLYLeuAspArgAlaTyrHisAspThrTyrTyrValValAlaHisPhe 411
|||||
491 TCCTCATGACGTTGTTTCATGATACCTACTAGCTGAGTACTGACTT 540
|||||
411 eHisTyrValMetSerLeuGlyAlaIlePheAlaIlePheAlaGlyIleT 428
|||||
541 TCACATGACGTTCTTCATGATGAGGCGCTGATTTGCAATCTGCGGTGTC 590
|||||
428 YrPheTyrMetProLysPheSerGlyYarGlyAlaPheProGluTrrPAlaAla 444
|||||
591 CTCACATGTTCCCTCTCTCTGCTGATATAGCTACACCCATATAGAGA 640
|||||
445 LysLeuHisPheTrpThrPhePheIleGlyAlaAsnValThrPhePhe 461
|||||
641 AAGGTTCACTTCTTCATATGTTGTTGAGTCACTTACCTTTTTC 690
|||||
461 oGlnHisPheLeuGlyYarGlnGlyMetProArgArgTyrIleAspTyr 478
|||||
691 TCACACTTCTTANGTCTACGCGGATGCCAGACGCTACTCAGACTATC 740
|||||
478 roGluAlaPheAlaLeuTrrPAsnLysValSerSerTyrGlyAla 492
|||||
741 CAGACGCTATACACTTTGAAATCTATCTNCATTCATTTGATGATCA 784
|||||
seq_name: gb-est2:BI101730

seq_documentation_block:
LOCUS BI101730 1007 bp mRNA EST 26-JUN-2001
DEFINITION 602887560F1 NCI_CGAP_Kid14 Mus musculus CDNA clone IMAGE:5042751
VERSION BI101730.1 GI:14552623
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS NIH-MGC http://mgc.nhl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@ds-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1M11117 row: h column: 16
High quality sequence stop: 840.
Location/Qualifiers
1..1007
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5042751"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-Sport6; Site_1: NCI;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT 286 a 251 c 161 g 309 t
ORIGIN

```

```

alignment_scores:
Quality: 861.50 Length: 306
Ratio: 3.405 Gaps: 5
Percent Similarity: 82.680 Percent Identity: 57.190

alignment_block:
US-09-712-768-2 x BI101730 ..
Align seg 1/1 to: BI101730 from: 1 to: 1007

193 ITeMeGlyAlaIleAsnMetIleThrThrPheLeuAsnMetArgAlaPr 209
|||||
1 ATTTAGGTGCAATTAAATTATTTATACCATATATATCAACATGAAACCC 50
209 oGlyMetThrLeuHisLysValProLeuPheSerTrpSerIlePheIleT 226
|||||
51 AGCCATACACAGTATCAACATCCACTATTGTCGATCCGACTTATTA 100
226 hrAlaTrpLeuIleLeuLeuAlaLeuProValIleuAlaGlyAlaIleThr 242
|||||
101 CAGCGGTACTGCTCCTATATACATACAGAGTGTAGCGGACGATTA 150
243 MetLeuLeuThrAspArgAsnPheGlyThrThrPhePheAsnProAlaG 259
|||||
151 ATACTACTAACAGACCGCAACCTTAACAACTTTCTTTGATCCCGCTG 200
259 yGlyGlyAspProIleLeuTyrGlnHisIleLeuThrPhePheGlyHisP 276
|||||
201 AGGAGGGACCAATCTCTACACAGATCTGTTGATTTCTTTGGGACC 250
276 roGluValTyrIleIleIleLeuProGlyPheGlyIleIleSerHisVal 292
|||||
251 CAGAGTTATATCTTATATCTCCAGAGATTTGAAATATTTACATGTA 300
293 ValSerThrPheSer...LysLysProValPheGlyTyrLeuProMetVal 308
|||||
301 GTTACTTACTACTCGGAATAAAGAACCTTTCGGCTTATATGAAATGT 350
308 IYrAlaMetValAlaIleGlyValLeuGlyPheValAlaTrrAlaHis 325
|||||
351 ATAGGAATATATGTCATTTGGCTTCTAGGCTTATTTGATAGAGCCACC 400
325 IAsMetGlyThrValGlyMetSerLeuThrGlnGlySerTyrPheMetLeu 341
|||||
401 ACATATTCACAGTAGAGATAGATGATGACACACAGAGCTTACTTACATCA 450
342 AlaThrMetValIleAlaValProThrGlyIleLysIlePheSerTrpI 358
|||||
451 GCCACTATATATTCGCAATTCCTACCGGTGTCAAGATATTTAGCTGACT 500
358 eAlaThrMetTrpGlyGly...SerValGlyPheLysSerProMetLeuTrp 374
|||||
501 TGCACACCTTACACGAGGATACCTATTAATGATCTCCAGCTATACATGA 550
375 AlaPheGlyPheMetSerPheLeuPheThrValGlyGlyValThrGlyIleVal 391
|||||
551 GCGTTAGGCTTATTTCTTATTATACAGTGTGGTGTCAACCGGAATGCT 600
391 ILeuAlaGlnAlaGlyLeuAspArgAlaTyrHisAspThrTyrTyrVal. 407
|||||
601 TTATTCACACATCATCCCTTGACATGCTGCTTACGATCATATCATATGTTA 650
408 ValAlaHisPheHisTyrValMetSerLeuGlyAlaIlePheAlaIlePhe 424
|||||
651 GTAGCCCATTTCCACATGTTCTATCAATAGGAGGACGATTTGCTATCAT 700
424 eAlaGlyIleTyrPheTyrMetProLys...PheSerGlyAlaGlyAlaPhe 440
|||||
701 AGGAGATTTGTTCACTGATTCGCCATATTTTCAAGGCTTACCCCTAGA 750
440 oGluTrrAlaIleAlaLysLeuHisPheTrpThrPhePheIleGlyAlaAsn 457
|||||
751 TGCACATTTGACAAAAGCCACTTGCATCATATTTCTGATAGGAATGAACA 800

```



```

457 a1rhrpheeprgln.hisphleuglyarglnclymetrproargar 473
      :|||||
801 TAACATTCCTCCCAACATTCCTCCGCGCTTACAGAAATACCCGAGC 850
      :|||||
473 gtyrilleasprglnalaphalealeutrpasnlysvalsest 490
      :|||||
851 CTACAGATACAGATGCTTACACACATTAACACTGCGCTCCCTA 900
      :|||||
490 yrcglyalphe 493
      :|||||
901 TAGGACATTTT 911
      :|||||

seq_name: gb_est1:AUI69837

seq_documentation_block:
LOCUS AUI69837 809 bp mRNA EST 29-JAN-2001
DEFINITION AUI69837 OL-br-ad cDNA Oryzias latipes cDNA clone br5288, mRNA
sequence.
ACCESSION AUI69837
VERSION AUI69837.1 GI:12591906
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE
1 (bases 1 to 809)
Mita,K., Ishikawa,Y. and Yamauchi,M.
Establishment of cDNA database of medaka, Oryzias latipes
unpublished (2001)
JOURNAL Contact: Mita K
COMMENT Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmk@nirs.go.jp
method:uni-directional sequence direction:sequenced from T3 primer
(5' -> 3').

FEATURES
source 1..809
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="br5288"
/clone_lib="OL-br-ad cDNA"
/sex="Female/male mixed"
/tissue_type="brain"
/dev_stage="adult"

BASE COUNT 203 a 193 c 153 g 259 t 1 others
ORIGIN

alignment_scores:
Quality: 858.50 Length: 265
Ratio: 3.700 Gaps: 1
Percent Similarity: 87.547 Percent Identity: 61.887

alignment_block:
US-09-712-768-2 x AUI69837 ..

Align seg 1/1 to: AUI69837 from: 1 to: 809

222 Serilepheeprglnalaphalealeutrpasnlysvalsest 238
      :|||||
11 GCCGCTACTATCACCGCAGATTTACTTCTTCTCTCCCTGTGCTAGC 60
      :|||||
238 aglyalaleprmetleuleutrpasnlysvalsest 255
      :|||||
61 TGCAGGATACATATCTTACAGACCGAAACCTAAATACACCTTCT 110
      :|||||
255 heasnproalaglyalaphalealeutrpasnlysvalsest 271
      :|||||

```

```

111 TCACCCCTGCAGAGAGAGGAGCCCTATTCCTACCAACCTTTCGTA 160
272 Phehecllyhsproglyvaltyrilleleleuleuproglypheglyl 288
      :|||||
161 TTTCTTGGCCACCCCTGAAGTTTATTTCTATTCCTCCCGCGCTGAAT 210
      :|||||
288 eileserhivalvalserthrpheser...lyslsprovalpheglyl 304
      :|||||
211 GATTTCTCATATGTGGCTTATTTATTCAGGTAAAGAACCTTTCGGCT 260
      :|||||
304 ylleuprometvaltyralametyalaleglyvalleuglypheval 320
      :|||||
261 ACATGGGAATGTATGAGCCATGATGCTATCGGACTTCTAGGTTATT 310
      :|||||
321 Valtrpalehshismetlyrthrvalglymetserleuthringle 337
      :|||||
311 GTTGGACTCACCATATTCACCGTAGGATGACGTGGACACTCGAGC 360
      :|||||
337 rtyrphemetleualathmetvalillealavalprothglylleysl 354
      :|||||
361 TTATTTTACATCTGCACACATATTCATTCGATTCCTAGCGGTTAAG 410
      :|||||
354 lephesertrpillealathmetrpglylservalgluphelysser 370
      :|||||
411 TATTTAGCTGACTGCTACCTGCTGATGAGTTCAATTAATGAGAAAC 460
      :|||||
371 prometleutrpalephglyphemetpheleupherthrvalglyl 387
      :|||||
461 CCATATTATGAGGCGCTGTTATTTCTCTTTCACGTGGGGGATCT 510
      :|||||
387 lthrglyllevalleualaglylleasprglnalaphalealeutrp 404
      :|||||
511 AACTGGCATGCTGTACCAACATCATCTAGATATATCTGATGATACA 560
      :|||||
404 htyrtyrvalvalalalehshismetlyrvalmetserleuglyal 420
      :|||||
561 CTACTATGTAGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610
      :|||||
421 Phalalephehaleaglylletyrphetyrmetprolysheserly 437
      :|||||
611 TTGCGCTTATGAGGCGCTTGTCTGCTGCTGCTGCTGCTGCTGCTG 660
      :|||||
437 galapheprglnalaphalealeutrpasnlysvalsest 454
      :|||||
661 TACTTTACAGGAAACATGAGCAAAAGATCCTGCGGTATATTTGCTG 710
      :|||||
454 lyalalevalthrphemetleuleutrpasnlysvalsest 470
      :|||||
711 GGGTAAATTTTAACTTCTTCTTCAACACTTCTTACGCGCTAGCGGA 759
      :|||||
471 Proargatrtyleasprglnalaphalealeutrpasn 485
      :|||||
760 CCACGACGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 804
      :|||||

seq_name: gb_est1:BE468572

seq_documentation_block:
LOCUS BE468572 911 bp mRNA EST 28-JUL-2000
DEFINITION Iphak00346 Head kidney cDNA library Ictalurus punctatus cDNA
similar to Cytochrome c oxidase subunit I, mRNA sequence.
ACCESSION BE468572
VERSION BE468572.1 GI:9559063
KEYWORDS EST.
SOURCE channel catfish.
ORGANISM Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Siluriformes; Ictaluridae; Ictalurus.
REFERENCE
1 (bases 1 to 911)
Cao,D., Kocabas,A., Ju,Z., Karsi,A., Li,P., Patterson,A. and Liu
Z.J.
Transcriptome analysis of channel catfish (Ictalurus punctatus):
Gene catalogue and profiling from the head kidney

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 10:54:55 ; Search time 39.94 Seconds
(without alignments)
1033.020 Million cell updates/sec

Title: US-09-712-768-2

Perfect score: 3000
Sequence: 1 MADAAIHGHHEKGFETR.....PAHTFTLPKRSWDKHPHS 557

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_1101.*
1: /SID88/gcgdata/geneSeq/geneSeq/AA1980.DAT.*
2: /SID88/gcgdata/geneSeq/geneSeq/AA1981.DAT.*
3: /SID88/gcgdata/geneSeq/geneSeq/AA1982.DAT.*
4: /SID88/gcgdata/geneSeq/geneSeq/AA1983.DAT.*
5: /SID88/gcgdata/geneSeq/geneSeq/AA1984.DAT.*
6: /SID88/gcgdata/geneSeq/geneSeq/AA1985.DAT.*
7: /SID88/gcgdata/geneSeq/geneSeq/AA1986.DAT.*
8: /SID88/gcgdata/geneSeq/geneSeq/AA1987.DAT.*
9: /SID88/gcgdata/geneSeq/geneSeq/AA1988.DAT.*
10: /SID88/gcgdata/geneSeq/geneSeq/AA1989.DAT.*
11: /SID88/gcgdata/geneSeq/geneSeq/AA1990.DAT.*
12: /SID88/gcgdata/geneSeq/geneSeq/AA1991.DAT.*
13: /SID88/gcgdata/geneSeq/geneSeq/AA1992.DAT.*
14: /SID88/gcgdata/geneSeq/geneSeq/AA1993.DAT.*
15: /SID88/gcgdata/geneSeq/geneSeq/AA1994.DAT.*
16: /SID88/gcgdata/geneSeq/geneSeq/AA1995.DAT.*
17: /SID88/gcgdata/geneSeq/geneSeq/AA1996.DAT.*
18: /SID88/gcgdata/geneSeq/geneSeq/AA1997.DAT.*
19: /SID88/gcgdata/geneSeq/geneSeq/AA1998.DAT.*
20: /SID88/gcgdata/geneSeq/geneSeq/AA1999.DAT.*
21: /SID88/gcgdata/geneSeq/geneSeq/AA2000.DAT.*
22: /SID88/gcgdata/geneSeq/geneSeq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3000	100.0	557	22	AAV97750
2	1169.5	39.0	581	22	AAAG92515
3	1160	38.7	552	22	AAAB79569
4	1078.5	36.0	648	22	AAAG81983
5	1010.5	33.7	663	22	AAAG98862
6	552.5	18.4	283	22	AAAB79571
7	340.5	11.3	204	22	AAAB79572
8	333.5	11.1	198	21	AAAB41481
9	303	10.1	56	22	AAV97762
10	186	6.2	472	21	AAAB26913
11	171	5.7	375	18	AAW20731

12	145.5	4.8	38	21	AAAB7168	Human prostate can
13	145.5	4.8	38	21	AAAB44050	Human cancer assoc
14	136	4.5	95	22	AAAB9570	Corynebacterium g1
15	136	4.5	377	22	AAU03645	Group B Streptococ
16	130.5	4.3	543	22	AAAG91478	C glutamicum prote
17	125.5	4.2	660	21	AAV74602	Neisseria meningit
18	123	4.1	1019	22	AAAG92745	C glutamicum prote
19	121.5	4.0	418	22	AAAG91830	C glutamicum prote
20	120.5	4.0	660	21	AAV74601	Neisseria meningit
21	116.5	3.9	1962	20	AAV17250	Namg polypeptide
22	115.5	3.9	386	22	AAAB79926	Corynebacterium g1
23	115	3.8	492	12	AAAB11360	Glucose Transporte
24	113	3.8	559	21	AAAB15974	E. coli proliferat
25	112.5	3.8	436	22	AAAG90123	C glutamicum prote
26	111	3.7	337	22	AAAG92904	C glutamicum prote
27	109.5	3.6	595	22	AAAG93189	C glutamicum prote
28	109	3.6	492	18	AAAB77835	Human glucose tran
29	108	3.6	452	22	AAAG93214	C glutamicum prote
30	108	3.6	452	22	AAAB76708	Corynebacterium g1
31	108	3.6	492	17	AAAB93000	Human GluT1 protei
32	107.5	3.6	484	22	AAAB76735	Corynebacterium g1
33	107	3.6	604	21	AAV75522	Neisseria meningit
34	106.5	3.5	384	20	AAV38852	Neisseria meningit
35	106.5	3.5	512	20	AAV38556	Neisseria meningit
36	106.5	3.5	660	21	AAV74600	Neisseria gonorrhe
37	106.5	3.5	722	21	AAV6163	Arabidopsis thalia
38	106	3.5	442	19	AAAB98870	H. pylori GHPo 171
39	106	3.5	442	19	AAAB71557	Helicobacter poly
40	106	3.5	442	22	AAAB6375	H. pylori HPN091 P
41	106	3.5	534	20	AAV72788	Glucose transporte
42	106	3.5	617	22	AAAB84532	Amino acid sequenc
43	106	3.5	617	22	AAAB84534	Amino acid sequenc
44	106	3.5	766	21	AAV6161	Arabidopsis thalia
45	105	3.5	431	22	AAAG82736	S. epidermidis ope

ALIGNMENTS

RESULT 1	
AAV97750	
ID AAV97750 standard; Protein; 557 AA.	
XX	
AC AAV97750;	
XX	
DT 06-AUG-2001 (first entry)	
XX	
DE G. oxydans cytochrome C oxidase (COI) protein sequence.	
XX	
KW Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;	
KW oxidative fermentation; electron transfer; respiratory chain; L-sorbose;	
KW 2-Keto-L-gluconic acid; 2KGA production; aldehyde production;	
KW carboxylic acid production; ketone production.	
XX	
OS Gluconobacter oxydans.	
XX	
PM EPI1103603-A2.	
PN	
PD 30-MAY-2001.	
XX	
PF 14-NOV-2000; 2000EP-0124785.	
XX	
PR 17-NOV-1999; 99EP-0122842.	
XX	
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.	
XX	
PI Asakura A, Hoshino T, Shinjoh M;	
XX	
DR WPI: 2001-357953/38.	
XX	
DR N-PSDB: AAA91489.	
XX	
PT New cytochrome c oxidase complex having cytochrome c oxidase activity	
PT from Gluconobacter oxydans DSM 4025, useful in mediating electron	

QY	308	VYAMAIVGLGVLEAVNNHNYTGMSLTQGSYFMLATMTVAIPGKIKFSWIATMMGGSVE	367
Db	295	IlaLtlSgdlSmavahhmVtgvall--pLtsfmflIsVpCyKfInwvctmWgHt	352
QY	368	FKSPMLMAFGFELFTVGCVTGIVLAQACIDRAVHDYTVVVAHFHYMSLGAIFAFACI	427
Db	353	wetpmltwsvygmattflfglgtlgimasppdlhdadsyIahbfytlgtvfvfascav	412
QY	428	YVYMKFSGRAPEYMAAKIHFTFTFICGAVNTFFPQHFLGRGSMFRXYDY--PEAFALIN	485
Db	413	yLwFpmtctgimmdetLgKhfwlftvgfhtllqhwvnmmprryadylsdgftIyn	472
QY	486	KVSSGAFIAFASFLEFIYFVTVVAGRERTRNPNGEFAADLEWTLSPPAHFTETL	545
Db	473	qLstvfslfIglstsvpfInnvfkswrgelvtvdwdpwg-ygnslwatscprphnfasl	531
QY	546	PK---RSDWDKNPSH	557
Db	532	pLrserpafelhYph	547
RESULT 3			
AAB79569			
ID	AAB79569 standard; Protein; 552 AA.		
XX	AAB79569;		
AC	AAB79569;		
XX	30-APR-2001 (first entry)		
DT	Corynebacterium glutamicum SMP protein sequence SEQ ID NO:654.		
DE	Corynebacterium glutamicum; carbon metabolism and energy production;		
XX	SMP protein; sugar metabolism and oxidative phosphorylation protein;		
KW	fine chemical production; organic acid; pyrimidine base; nucleoside;		
KW	nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;		
KW	nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;		
KW	carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;		
KW	diagnosis; Corynebacterium diptheriae; evolutionary study.		
XX	Corynebacterium glutamicum.		
OS	MO200100844-A2.		
XX	04-JAN-2001.		
PN	23-JUN-2000; 2000MO-IB00943.		
XX	25-JUN-1999; 99US-0141031.		
PR	08-JUL-1999; 99DE-1031412.		
PR	08-JUL-1999; 99DE-1031413.		
PR	08-JUL-1999; 99DE-1031419.		
PR	08-JUL-1999; 99DE-1031420.		
PR	08-JUL-1999; 99DE-1031424.		
PR	08-JUL-1999; 99DE-1031428.		
PR	08-JUL-1999; 99DE-1031431.		
PR	08-JUL-1999; 99DE-1031433.		
PR	08-JUL-1999; 99DE-1031434.		
PR	08-JUL-1999; 99DE-1031510.		
PR	08-JUL-1999; 99DE-1031562.		
PR	08-JUL-1999; 99DE-1031634.		
PR	09-JUL-1999; 99DE-1032180.		
PR	09-JUL-1999; 99DE-1032227.		
PR	09-JUL-1999; 99DE-1033230.		
PR	09-JUL-1999; 99US-0145208.		
PR	14-JUL-1999; 99DE-1032924.		
PR	14-JUL-1999; 99DE-1032973.		
PR	27-AUG-1999; 99DE-1033005.		
PR	31-AUG-1999; 99US-0151572.		
PR	03-SEP-1999; 99DE-1042076.		
PR	03-SEP-1999; 99DE-1042079.		
PR	03-SEP-1999; 99DE-1042086.		
PR	03-SEP-1999; 99DE-1042087.		

[illegible]

OY 442 WAAKTHHEWTFEFGANVTPPEQHFELGROGMPPRYIDY--P6AFALMNKXSSGAFLAFAASF 499
Db | - ||| : : : : || :||::| ||||| || : :||::| : |||
 398 rlgkhhfwlcltvgifngftligwvnmqmprryadyldsdgfltnqlstlvfyllglstv 457

OY 500 LEFFIVFYMTLVAGRRERTRPNMCFADLTLEWTLPSPPAHTFEETLPR----RSDMDKH 555
 : || : : : : ||| : :||| ||| : :||| : :||| : :|||
Db 458 lpfimvfekswrygelvtvdwpwg-ynsllewatscprrhnfaelprlseerpaefehly 516

OY 556 SH 557
 |
Db 517 ph 518

RESULT 4
AAG81983
ID AAG81983 standard; Protein: 648 AA.
XX
AC AAG81983;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:1060.
XX
KM Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX vaccination; endocarditis.
OS Staphylococcus epidermidis.
PN WO200134809-A2.
PD 17-MAY-2001.
PF 09-NOV-2000; 2000WO-US30782.
PR 09-NOV-1999; 99US-O164258.
PS (GLAX) GLAXO GROUP LTD.
PT kimmerly WJ;
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
DR useful for vaccinating against infections, e.g. endocarditis -
N-PSDB; AAH52833.

Claim 18; Page 309; 2188pp; English.
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
(II), given in AAG81454 to AAG83120, from staphylococcus epidermidis.
(I) and (II) can have antibacterial activity and therefore can be used
in vaccination. The nucleic acids (I) may be used to produce the
S. epidermidis polypeptides (II) via the production of vectors
containing them which are used to produce hosts cells which express the
polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
used to vaccinate subjects and to raise antibodies against the bacteria
CC The polypeptides may also be used to assay for other inhibitors of their
activity and therefore identify compounds that may be used for the
treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
AAH55090 represent specifically claimed S. epidermidis genomic DNA
polynucleotide sequences from the present invention. AAH55091 to
AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
sequences given in the sequence listing of the present specification,
however the sequence listing only goes up to SEQ ID NO:4454 so even
though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
no sequences are present for SEQ ID NO:4455 to 4464.

Sequence 648 AA;

Query match	36.0%; Score 1078.5; DB 22; Length 648;
-------------	---

	Best Local Similarity	38.5%: Pred. No. 8, 2e-100:	
	Matches	212; Conservative 10%; Mismatches 203; Indels 33; Gaps	8:
QY	18	ETRWENSTNHHKDIGLLIYAAGVGEISVLTFTVMRLELMDRGVQMYMCEGARLIADASQ	77
DB	30	YKWFETSVCHKKIGIMYLISAVLMFVGGIDALMLIRTLQTLIPDNKFL-----	76
QY	78	TCYANGHLENWVYVYGIIMFVFGIPALFEGGNGVLMLOIGADPMAPRRNNISPMLEF	137
DB	77	-----eanhynevtcthyvmlilfmaprlf-glmvnpdlqgadvatprvwnsvfwlf	131
QY	138	IAGTAMGVASLEPAGDGGQLGSGVGWVLYPLSTR-EAGYSMDLAIFAVHLSGASSINGA	196
DB	132	fagmlifnlis-fivvg-----spaagwtuyaprlagetppgyvnylliaqlsgjsglmtg	186
QY	197	INMITEFLMKRAPGMTLHKVPLFNSISIFTNALLDLAPVLAGAITMLTDNRNGETTFPN	256
DB	187	lffvfllyckrptlckfmgpmfsvtctfcltlivlalfvrfvalaImbadrlfgtqfft	246
QY	257	PAGGGDPILYOHLMEFPGAREVYIIILPEFGILSHVSYFESKKPYEGVLPVYAWVAIGV	316
DB	247	vandgmpmlwanfwwfwgipreyivlrlpafgmyselirlfakrlfngsmIwataglaf	306
QY	317	LGFWVAHHMHTVGSLSLTOOSTYFMLATFVIAVPTGIKIFSWIATMGGSVFEKSPMLAF	376
DB	307	lsfllyvwhfflmngallnsfslstmltgyrvtgvyklfnwlltlykgqltfespmlesl	366
QY	377	GMFLFTVGGVGIYLAQGLDRAVHDPRVYVAHHHYWMSLGAITAIAGIYFYMPKFSG	436
DB	367	afimnfllygvcgvmIamasadyhncyflvahthryltvrgvfacIagllfwypkmmg	426
QY	437	RAEPMAAKLMEHFEFICGANVTFPOHFLGROGMRRRIDYPEAFALV--NKVSSYGAFI	494
DB	427	ylkneclnkwcftwflmnglnvcflpqflgltdgmprrltylmpsdgwwllnftslcIagvl	486
QY	495	AASAPLFLFIVIVYVLYLVAGRRETRPNMGCEPADTLEMTLPSP--PRAHFETLPRKSDMD-	552
DB	487	maigflflvasivshikapreatgdnwdgIgrtllewsIasalppkynfaltpdwndydt	546
QY	553	-----KHPSH 557	
DB	547	fydmkkehgrh 556	
RESULT	5		
AA98862	ID	AA98862 standard; Protein; 663 AA.	
XX	AC	AA98862;	
XX	DT	26-SEP-2001 (first entry)	
DE	XX	E. coli growth and proliferation related protein sequence SEQ ID NO:332.	
XX	XX	Escherichia coli; growth; proliferation; microbial; antimicrobial;	
KW	XX	bacterial infection; microorganism.	
OS	XX	Escherichia coli.	
XX	PN	MO200134810-A2.	
XX	PD	17-MAY-2001.	
PF	XX	09-NOV-2000; 2000MO-US30950.	
PR	XX	09-NOV-1999; 99US-0164415.	
PA	XX	(ELIT-) ELITRA PHARM INC.	
XX	PI	Forsyth RA, Ohlsen K, Zyskind J;	
XX	XX	WPI; 2001-335933/35.	
DR	XX	N-PSDB; AAH84533.	

PS Claim 20; Page 1076-1077; 1246pp; English.

XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein sequences required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).

XX Sequence 283 AA:

SO

Query Match 18.4%; Score 552.5; DB 22; Length 283;
 Best Local Similarity 41.3%; Pred. No. 2.4e-47;
 Matches 104; Conservative 47; Mismatches 92; Indels 9; Gaps 4;

OY 312 VAGGVGLFVVAHMHVYVMSLQGSYFMLATVIAVPGIKIFESMTATMGSSVEKSP 371
 Db 1 Isigalsmawahimltvtgavll--pfisfmtilsvrpgvkltnvgtmwkhltwep 58
 OY 372 MLNAFGFMFLFTVGTGVLALQADRAYHDYVVAHNFHYVMSLGAIFAFAGTIFYM 431
 Db 59 mltvsgfmatflfgslgmlasprldfnhdsyfliahbthlyftvtvfasagtyvlf 118
 OY 432 PKRSGRAFRPMANKLHFWTFITGANVTFEPQHLGROGMPRTIDY--PEAFYLMKVVSS 489
 Db 119 phtgtgmmrdlghkthwltfvgfhgtflfghvgnmgmptryadyldsdgftlyngist 178
 OY 490 YGAFLEAFSLFFVTVFVTVLVAGRRTRPNMGCEFADTLEWLPSPPAHPEETLPK-- 547
 Db 179 vtflllgslvrlfwmvfkswtygelvtvdwpg-ygnslwatsecpplmhtasiprlir 237
 OY 548 --RSDMDKHPSH 557
 Db 238 serpafelhyph 249

RESULT 7
 AAB79572 ID AAB79572 standard: Protein; 204 AA.
 XX AAB79572;
 XX 30-APR-2001 (first entry)
 XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:660.
 XX
 XX Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX Corynebacterium glutamicum.
 OS
 XX
 XX WO200100844-A2.
 PN
 XX
 PD 04-JAN-2001.
 XX

PF 23-JUN-2000; 2000WO-IB00943.
 XX
 PR 25-JUN-1999; 990S-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042099.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.

PA (BADI) BASF AG.
 XX
 PI Pompejus M. Kroege B. Schroeder H. Zelder O. Haberbauer G.
 XX
 DR WPI; 2001-061975/07.
 N-PSDB; AAF71689.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 1079; 1246pp; English.

CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein sequences required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).

XX
 XX Sequence 204 AA:

SO

Query Match 11.3%; Score 340.5; DB 22; Length 204;
 Best Local Similarity 41.6%; Pred. No. 3.4e-26;
 Matches 67; Conservative 31; Mismatches 56; Indels 7; Gaps 3;

XX Example 3; Page 28; 42pp; English.

CC This sequence represents a fragment of the Gluconobacter oxydans
CC cytochrome C oxidase, COI. The invention relates to a cytochrome c
CC oxidase COI complex. The COI complex is useful in improving oxidative
CC fermentation and is an essential component mediating electron transfer
CC in the respiratory chain. The recombinant microorganism and the
CC cytochrome c oxidase may be used in the genetic preparation of a
CC recombinant COI complex and in the production of 2-Keto-L-gluconic acid
CC (2KG) from L-sorbose or D-sorbitol in a culture medium. The COI is also
CC useful as a terminal oxidase, oxidizing cytochrome c, an electron
CC acceptor from an enzyme belonging to dehydrogenase c, for the production of
CC aldehydes, carboxylic acids and ketones from alcohols and aldehydes,
CC especially the production of 2KG from L-sorbose or D-sorbitol.

XX Sequence 56 AA;

Query Match 10.1%; Score 303; DB 22; Length 56;
Best Local Similarity 98.2%; Pred. No. 3.8e-23;
Matches 55; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 271 WFGHPEVYIIILPGFGIISHVSTFSKRPVGYLPMYAMAIGVLGFVVAHHM 326
DB 1 wffghpevyiilipgfgiishvstfskrrpvfgylpmyamaigvlgfvvahn 56

RESULT 10

AAB26913 AAB26913 standard; Protein; 472 AA.

AC AAB26913;

DT 12-JAN-2001 (first entry)

XX Large subunit of nitric oxide reductase NorB.

XX NorB; nitric oxide reductase; enzyme; transgenic organism;

KM toxic waste denitrification; denitrifying bacteria.

XX Pseudomonas sp. G-179.

OS US6136588-A.

PN 24-OCT-2000.

XX 15-JUL-1999; 99US-0354129.

PR 17-JUL-1998; 98US-0093191.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Bedzyk LA, Ye RW;

PI WPI: 2000-686079/67.

DR N-PSDB; AAA94683.

XX New nucleic acid encoding bacterial nitric oxide reductase, useful e.g.
PT for preparing transgenic organisms able to detoxify wastes

XX Claim 1; Columns 21-24; 24pp; English.

CC The present sequence is the NorB protein from Pseudomonas sp. strain
CC G-179. This sequence is the large subunit of nitric oxide reductase. The
CC NorB gene can be used to create transgenic organisms that can denitrify
CC toxic waste. In addition, the NorB gene may be used to produce probes and
CC primers for identification of new denitrifying bacteria.

XX Sequence 472 AA;

Query Match 6.2%; Score 186; DB 21; Length 472;

Best Local Similarity 21.7%; Pred. No. 3.6e-10;
Matches 98; Conservative 77; Mismatches 173; Indels 104; Gaps 23;

OY 89 MWTYGIIMFFVFGICPALFGCGEYNTLMPLOIGAPDPAEPKRNUNLSFWFLAGTGMGASL 148

DB 51 mhtalilwlllg----fmgaayllp-eeaeetlyspklalagfwlflaaalavv- 103

OY 149 FAPGDDGLGSGVGVLPPLSTREAGYSMDLAFVAHLGASSIMGAINMTTFELMRA 208

DB 104 -----gymfkinegreffleqfllkiglviv-----cimflnvtmstlgrtk 146

OY 209 PGMTHKVPLES-MSIFITAMLLIALPVLGATMTLLTDNFGTTFEPNAG-OGDEPLY 266

DB 147 tvvt-nlllglwglalf-----flsfypnaldlkmw 181

OY 267 OHT--LMFFGHPEVYIIILPGF-----GIISHVSTFSKRPVGYLPMYAMAIGV 314

DB 182 yvvvhlwvegwelmasvllaflmklngldrevvkw-----lyvlgglalfs 230

OY 315 GVLGFVVAHHMYTIVGM-SLNOQSYFMLATWVIAVPTGIRKIFSMATMW--GGSVEKSP 371

DB 231 gllg---tghlyywigapgywqwisltlevapfflmvflftvmt-wkagtrhpnraa 286

OY 372 MLMAFG-FMFLFTVGVGTIVLAQAGLDRAVHDITYVVAHMF-----YVNSLGATPAIF 424

DB 287 llwstgsvmafifgagvwgflhtlssvnylchgcvtahghlaiffgaymnlalmaa-- 344

OY 425 AGIYFYMRFSGR-AFPWMAKLHFWTFEGANVTFFQHLG-----RQGMRRRTIDY 477

DB 345 ---yavpelcgrkpynglsmvsfmmctamsvmltalafagylqvlhqlvlggydv 400

OY 478 PEAFAL--MNKVSYGAFLAFAFLFTVIVFV 507

DB 401 qqlalifywrlgs-gvflvllsalmtfwallv 431

RESULT 11

AAM20731 AAM20731 standard; protein; 375 AA.

AC AAM20731;

DT 16-JUL-1997 (first entry)

XX H. pylori inner membrane protein, 06cpl118orf6.

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

XX Helicobacter pylori.

OS WO9640893-A1.

PN 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

PR 01-APR-1996; 96US-0630405.

PR 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

PI WPI: 1997-052306/05.

DR N-PSDB; AAT67984.

XX Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter

XX Claim 56; Page 1146-1147; 1481pp; English.

XX The present sequence is a Helicobacter pylori inner membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed by ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORFs were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.

XX Sequence 375 AA:

Query Match 5.7%; Score 171; DB 18; Length 375;
Best Local Similarity 22.2%; Pred. No. 8.7e-09;
Matches 75; Conservative 64; Mismatches 143; Indels 56; Gaps 17;

QY 222 SIFITAWLILALPYLAGAITMLTRNGTFFNPAG-----GCDPIIYQHILW 271
DB 48 Ilyslwlyiat--yvglaavmyifnlsipcyfvdmgsvshsismysgndalig--w 102
QY 272 FFGHEVYIIILPG-FGLISHVSTFSKRPVGYLPMYANVAIGVGFV-WMA--HHMY 327
DB 103 wghnaavafvrlsgyilgyiflyfllkpsqpflyskltflfsws---lmfyilwagshhll 159
QY 328 TVGMSLTQGSFEMLATVIAVPTGKIFSWIATMWSGSVEFK-SPMLMAFGFMEF---F 382
DB 160 ystvdpdwqtlssvsvslilpswgtalimllmvgqwhqkpspli---kfvlslstly 216
QY 383 TVGVTGIVLAQAGUDRAHYDHYVVAHFHYVMSIGAI-FAIFAGIYFYMKR-----FS 435
DB 217 mlstlegisigaksvnaahetdwlghvhdgv-lywvgfcliasmymprlFKrelys 275
QY 436 GRAPEMAKALHEWTFEFGANVTFEPQHLG-ROGMNPR-----YIDPEAFAL 483
DB 276 gr-----lvdfqfwmrlgvlvlyfssmviagltgmmwrdvdygnlltygldtckalip 330
QY 484 WNKVSSYGAFIAPASFFFIIVFYVTLVAGRR-ETRPN 520
DB 331 yynlgyvglymlyfifllyfaynfmcltagkklerepn 368

RESULT 12

AAB57168
ID AAB57168 standard; Protein: 38 AA.

XX AAB57168;

XX 13-MAR-2001 (first entry)

DE Human prostate cancer antigen protein sequence SEQ ID NO:1746.

XX Human: prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular;
KW vulnereary; gastroinhtestinal; nephrotoxic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastroinhtestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.

OS Homo sapiens.

PN WO200055174-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05988.

XX

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM;

DR WPI: 2000-587513/55.
N-PSDB; AAF16371.

XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
PS Claim 11: Page 2216; 2338pp; English.

CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytoskeletal,
CC cardioactive, immunomodulatory, muscular, vulnereary, gastrointestinal,
CC nephrotoxic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastroinhtestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.

XX Sequence 38 AA;

Query Match 4.8%; Score 145.5; DB 21; Length 38;
Best Local Similarity 73.7%; Pred. No. 1.6e-07;
Matches 28; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 272 FFGHEVYIIILPGFGLISHVSTFS-KRPVFGYGLMV 308
DB 1 ffghevyiililpgfglshvstfsgkkrpvglygm 38

RESULT 13

AAB44050
ID AAB44050 standard; Protein: 38 AA.

XX AAB44050;

XX 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1495.

XX Human: cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytoskeletal; proliferative; vulnereary; immunomodulatory;
KW antidiabetic; antihistaminic; antitubercular; antitubercular;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; noctropic;
KW vasotropic; antipsoratic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.

OS Homo sapiens.

PN WO200055350-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05882.

XX 12-MAR-1999; 99US-0124270.


```

Db 2 FADMELESTNHHKIDIGTLLLEGAWAGVLTALSLIRAEIGQPC-----45
QY 77 QTCANGHLNVMVYTYGILMPEFVGIPALFEGFNGYLMPLDJGAPDMAFRNNLSFNL 136
Db 46 -NLGNHITVITVTAIAFPIFMPWIPMIGGFGNVLVLMIGAPDMAFRNNNSFWL 104
QY 137 FIATGAVASLAPAGDGLGSGVGVLYPLSTR--EAGYSMDLAIFAHLGASSIM 194
Db 105 LPPSLLLLASAMY-----EAGAGTGWTVYPLAGNSHFGASVDLITFSLHLAGVSSIL 159
QY 195 GAIMITTFILNMRAPGTLTKVPLFSISFITAMLLALPVLGATITMLTTRNGCTTF 254
Db 160 GAINFITITINMKPRAPTQYQTPLEFVMSVLTAVLLSLPVLGATITMLTTRNLTTF 219
QY 255 FNPAGGDPILYOHILMPEFGPEYIIILPGFIISHVSTFS-KKRFEGYCLMVYAMVA 313
Db 220 FDPAGGDPILYOHILMPEFGPEYIIILPGFISMISHIIVYSGKKEPEFGMGMVAMMS 279
QY 314 IGVLFVVAHMHVYVMSLTQOSYEMLAIVAPPTGIKIFSMIATMMGSEFEKSPML 373
Db 280 IGFLEFVMAHMHFTVGMVDVTRAYFTSATMIIAIPGVKVFSLMALTLSGNNKMSAVL 339
QY 374 WAFBFMLFTVGGVTVGLVLAQGLDRAYHDTYVVAHFHVMGALFAIFACITFYEMK 433
Db 340 WALGFIFLTVGGTLGVLANSSLDIVLHDTYVVAHFHVLMSGAFALMGGFIMHPEL 399
QY 434 FSGRAPEMAKLFHFTFEGANVTFFPOHLCROGMPRYIYDPEAFALMNVSSYGF 493
Db 400 FSGYTLDTQYAKIHFTIMFGVNLTFPPQHLGLSGMPRYSYDPPAYTTWNLLSSVGF 459
QY 494 LAFASLFETVIEVYTLVAGRRETRPNMGEFAD-----TLEWTLPSPPAPT 541
Db 460 ISLFAVMLMIFMI-----WEAFASKRKVLVVEBPSNMLEWLYGCPPEYHT 504
QY 542 FE 543
Db 505 FE 506

RESULT 2
US-09-354-129-2
; Sequence 2, Application US/09354129
; Patent No. 6136588
; GENERAL INFORMATION:
; APPLICANT: YE, RICK W.
; TITLE OF INVENTION: GENES ENCODING DENITRIFICATION REACTIONS
; FILE REFERENCE: CL-1255
; CURRENT APPLICATION NUMBER: US/09/354,129
; CURRENT FILING DATE: 1999-07-15
; EARLIER APPLICATION NUMBER: 60/093,181
; EARLIER FILING DATE: July 17, 1998
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Pseudomonas sp. strain G-179
US-09-354-129-2

Query Match 6.2%; Score 186; DB 4; Length 472;
Best Local Similarity 21.7%; Pred. No. 3.3e-10;
Matches 98; Conservative 77; Mismatches 173; Indels 104; Gaps 23;

```

```

Db 147 IYVT--NILIFGLMGLAIF-----FLFSFYNPANLADKKMY 181
QY 267 QH--LMFEGHEVYIIILPGF-----GILSHVSTFSKKRPFVCLPMVYAMVAI-----314
Db 182 WYVHLMEGVWELIMASVLAFLMKLNGIDREVEKW-----LYVIGLALES 230
QY 315 GVLGFVVAHMHVYVGM-SLTQOSYEMLATVIAVPTGIKIFSMIATMM--GGSVEFEKSP 371
Db 231 GILG---TGHYHWIGAPQYQWIGSLFSTLEVAFFFTMVFITFYMT--WKAGRKHPNRAA 286
QY 372 MLMAFG-EMELFTVGGVTVGLVLAQGLDRAYHDTYVVAHFH-----YVMSLGAIFAIF 424
Db 287 LMSIGCSYMAFFGAGVWGFLHTLSVNYTGTQVTAHGHLEFFGAYVMLNLAIMA-- 344
QY 425 AGIYFMPKFSGR-APPEMAKLFHFTFEGANVTFFPOHFLG-----RCGMPRYIDY 477
Db 345 ----YAVPEICGRKPNQWLSWVSTFMMCTANVTFALTFAGVLQVHLQVLRGEGYMDV 400
QY 478 PPAFAL--WNKVSYGAPLAFAPLFFVIFV 507
Db 401 QDQALFVYWRIGS--GVFLISALMFVAILV 431

RESULT 3
US-08-355-844-3
; Sequence 3, Application US/08355844
; Patent No. 5940307
; GENERAL INFORMATION:
; APPLICANT: Fischarg, Jorge
; APPLICANT: Czelegledy, Ferenc
; APPLICANT: Iserovich, Pavel
; APPLICANT: Li, Jun
; APPLICANT: Cheung, Min
; TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/355,844
; FILING DATE: 14-DEC-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Tang, Henry Y.S.
; REGISTRATION NUMBER: 29,705
; REFERENCE/DOCKET NUMBER: A29927-50/29910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2586
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: amino acid
; STRANDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Human
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..492
; OTHER INFORMATION: Facilitative glucose transporter

```

OTHER INFORMATION: Glut1 protein
US-08-355-844-3

Query Match 3.6%; Score 108; DB 2; Length 492;
Best Local Similarity 20.3%; Pred. No. 0.018;
Matches 92; Conservative 57; Mismatches 160; Indels 144; Gaps 19;

QY LMNMYV---TYHGILMFVIGIPALFEGFGNYLMPLO---IGAPDMAFPRMNN---LS 133
DB LMSLSVAIFSVGMIGSVGLFVNRFRGRNSMLMMLLAFVSAYLMGFSKLSFEMLI 123
QY 134 FWLFIAGTAMGVASLFPAPGSGDGLSGVWVLYPLSTREA---GYSMDLAFVHLS 188
DB 124 LGRFTIGVCGLTGTGFVPMYGEVS-----PTAFRGALGTLHQGLIVGILIAQVF 174
QY 189 GASSIMGAINMTTFLNMRAPGMTLHKVPLFSMSIFITAMLLILALPVLAGAITMLTDR 248
DB 175 GLDSTMGKDL-----WPLLSTIFIPALLOCIIVIPCPSPRFLINR 218
QY 249 N-----FGTT-----FNPAGGDPILVQHIL 270
DB 219 NEENRAKSVLKRLGTADVTHDLOEMKEESRQMREKKVTIILFRSPAYRQPIILAVVL 278
QY 271 WFGHPREVYIIILPGFGIISHVSTFSKRPVGYLPMYVAMVAGVLG-----FVY- 321
DB 279 QLSQO-----LSGINAVFYSTSIPEKA--GYOQPVYATIGSGIVMTAFVVSLEFVE 329
QY 322 ---WAHMYVGMISLQOSYFMLATVAVPTGIRKFSMTATMGSGVEKSPMLMAF 376
DB 330 RAGRTLHLIGLAGMA---GOAILMTIALAL--LEQLPMWSYL-----SIVAI 373
QY 377 GFMPLFTVG-----GVGTIVLAQAGLDRAVHDYVVAHFHYVMSLGA-- 419
DB 374 GVAFAFEVGPPIPMFTVIALESQGRPAALAVAGFSN--WTSNFIWGMCFQYVBDLCGPY 432
QY 420 IFALFAGI-----YFMPKFSGRAPEWMA 444
DB 433 VFILFTVLVLFIRTYFKVPETKGRTFDEIAS 465

RESULT 4
PCT-US95-16126-3
Sequence 3, Application PC/TUS95J16126
GENERAL INFORMATION:
APPLICANT: Fischberg, Jorge
APPLICANT: Czegledy, Ferenc
APPLICANT: Iserovich, Pavel
APPLICANT: Li, Jun
APPLICANT: Cheung, Min
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/355,844
FILING DATE: 14-DEC-1994
ATTORNEY/AGENT INFORMATION:

NAME: Tang, Henry Y.S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A29927-50/29910
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEO ID NO: 3
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..492
OTHER INFORMATION: Facilitative glucose transporter
OTHER INFORMATION: Glut1 protein
PCT-US95-16126-3

Query Match 3.6%; Score 108; DB 5; Length 492;
Best Local Similarity 20.3%; Pred. No. 0.018;
Matches 92; Conservative 57; Mismatches 160; Indels 144; Gaps 19;

QY 85 LMNMYV---TYHGILMFVIGIPALFEGFGNYLMPLO---IGAPDMAFPRMNN---LS 133
DB 64 LMSLSVAIFSVGMIGSVGLFVNRFRGRNSMLMMLLAFVSAYLMGFSKLSFEMLI 123
QY 134 FWLFIAGTAMGVASLFPAPGSGDGLSGVWVLYPLSTREA---GYSMDLAFVHLS 188
DB 124 LGRFTIGVCGLTGTGFVPMYGEVS-----PTAFRGALGTLHQGLIVGILIAQVF 174
QY 189 GASSIMGAINMTTFLNMRAPGMTLHKVPLFSMSIFITAMLLILALPVLAGAITMLTDR 248
DB 175 GLDSTMGKDL-----WPLLSTIFIPALLOCIIVIPCPSPRFLINR 218
QY 249 N-----FGTT-----FNPAGGDPILVQHIL 270
DB 219 NEENRAKSVLKRLGTADVTHDLOEMKEESRQMREKKVTIILFRSPAYRQPIILAVVL 278
QY 271 WFGHPREVYIIILPGFGIISHVSTFSKRPVGYLPMYVAMVAGVLG-----FVY- 321
DB 279 QLSQO-----LSGINAVFYSTSIPEKA--GYOQPVYATIGSGIVMTAFVVSLEFVE 329
QY 322 ---WAHMYVGMISLQOSYFMLATVAVPTGIRKFSMTATMGSGVEKSPMLMAF 376
DB 330 RAGRTLHLIGLAGMA---GOAILMTIALAL--LEQLPMWSYL-----SIVAI 373
QY 377 GFMPLFTVG-----GVGTIVLAQAGLDRAVHDYVVAHFHYVMSLGA-- 419
DB 374 GVAFAFEVGPPIPMFTVIALESQGRPAALAVAGFSN--WTSNFIWGMCFQYVBDLCGPY 432
QY 420 IFALFAGI-----YFMPKFSGRAPEWMA 444
DB 433 VFILFTVLVLFIRTYFKVPETKGRTFDEIAS 465

RESULT 5
US-09-031-392-4
Sequence 4, Application US/09031392
Patent No. 5942398
GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis A.
APPLICANT: Wang, Xun
TITLE OF INVENTION: NUCLEIC ACID MOLECULES
TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street

```

: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Melkijohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 534 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-031-392-4

```

```

Query Match 3.5%; Score 106; DB 2; Length 534;
Best Local Similarity 21.2%; Pred. No. 0.031;
Matches 98; Conservative 58; Mismatches 155; Indels 152; Gaps 22;

```

```

QY 40 VVGFTSVLTFTVYVRLMDPGVQVYMCLEGARL-----IADASQTCANGH 84
DB 142 VVNVLSIAGNLMGLAKMPS-HILTIAGRAITGLYCGLSGLVPMYSEVSPTA----- 195
QY 85 LNNVWVTVH-----GILMFPVGIPLAFGCGNYLMPLOIGAPDMAFPRMNNLSFWLFI 138
DB 196 LRGAIGTLHQLAIYTGILISQVGLDPLLG--NDELMPLLGLSGVA---ALLQFELL- 248
QY 139 AGTAMGVASLFAFGDQGLSGVGWVLYPLSTREAGYSMDLAIFAVHLSGASSIMGAIN 198
DB 249 -----LLCPESPXYLYIKLKV-----EAKKSLK-----RLRGNCDPMKEIA 286
QY 199 MITTFLNMRAPGMTLHKVPLFESMSIFITAMLILLAPVLAGATMILLDRN---FGTTF 254
DB 287 EMEKEKEAASEKRVISIGQLFSSSKYRQA--VIVALMV---QISQOFGINAIFYSTNI 341
QY 255 FNPAGGDPILYOHILMFEGHPEVYIIILPFGIISHVSTFS-----KKPVFGLLP 306
DB 342 FQRAVGQPPVYATL-----GCVGVNTVFYISVFLVEKAGRSLSF----- 382
QY 307 MYTAMVAICVLGFVVAHHMTYVGMSTLQOQSYFMLATWIAVPTGIKIFSWIATMGGSV 366
DB 383 -----LAGLMGMLISAAM-TVGLVLLSQ-----FAMMSYV----- 412
QY 367 EFKSPMLAEGFMFLFTVG-----GVTGIYLAQAGLDRAVHDITYVVAHF 411
DB 413 -----SWAIFLFLVFEVGPDPIMPFIVAELFSQGPRAAIAVAGFCN-WACNFIYGMCF 467
QY 412 HYVMSLGA--IFAIFAGI-----YFYMPKFSGRAPFEMA 444
DB 468 QYIADLCGPYVFVFAVLLVFLFAYLKVPETKGSFEEIAA 510

```

```

RESULT 6
US-09-299-549-4
: Sequence 4, Application US/09299549
: Patent No. 6136547
: GENERAL INFORMATION:
: APPLICANT: Tartaglia, Louis A.
: APPLICANT: Weng, Xun

```

```

: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
: TITLE OF INVENTION: GLUTEX AND USES THEREOF
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/299,549
: FILING DATE: 26-APR-1999
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Melkijohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 534 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-299-549-4

```

```

Query Match 3.5%; Score 106; DB 4; Length 534;
Best Local Similarity 21.2%; Pred. No. 0.031;
Matches 98; Conservative 58; Mismatches 155; Indels 152; Gaps 22;

```

```

QY 40 VVGFTSVLTFTVYVRLMDPGVQVYMCLEGARL-----IADASQTCANGH 84
DB 142 VVNVLSIAGNLMGLAKMPS-HILTIAGRAITGLYCGLSGLVPMYSEVSPTA----- 195
QY 85 LNNVWVTVH-----GILMFPVGIPLAFGCGNYLMPLOIGAPDMAFPRMNNLSFWLFI 138
DB 196 LRGAIGTLHQLAIYTGILISQVGLDPLLG--NDELMPLLGLSGVA---ALLQFELL- 248
QY 139 AGTAMGVASLFAFGDQGLSGVGWVLYPLSTREAGYSMDLAIFAVHLSGASSIMGAIN 198
DB 249 -----LLCPESPXYLYIKLKV-----EAKKSLK-----RLRGNCDPMKEIA 286
QY 199 MITTFLNMRAPGMTLHKVPLFESMSIFITAMLILLAPVLAGATMILLDRN---FGTTF 254
DB 287 EMEKEKEAASEKRVISIGQLFSSSKYRQA--VIVALMV---QISQOFGINAIFYSTNI 341
QY 255 FNPAGGDPILYOHILMFEGHPEVYIIILPFGIISHVSTFS-----KKPVFGLLP 306
DB 342 FQRAVGQPPVYATL-----GCVGVNTVFYISVFLVEKAGRSLSF----- 382
QY 307 MYTAMVAICVLGFVVAHHMTYVGMSTLQOQSYFMLATWIAVPTGIKIFSWIATMGGSV 366
DB 383 -----LAGLMGMLISAAM-TVGLVLLSQ-----FAMMSYV----- 412
QY 367 EFKSPMLAEGFMFLFTVG-----GVTGIYLAQAGLDRAVHDITYVVAHF 411
DB 413 -----SWAIFLFLVFEVGPDPIMPFIVAELFSQGPRAAIAVAGFCN-WACNFIYGMCF 467
QY 412 HYVMSLGA--IFAIFAGI-----YFYMPKFSGRAPFEMA 444
DB 468 QYIADLCGPYVFVFAVLLVFLFAYLKVPETKGSFEEIAA 510

```


Query Match 3.5%; Score 106; DB 1; Length 617;
 Best Local Similarity 19.8%; Pred. No. 0.038;
 Matches 69; Conservative 43; Mismatches 117; Indels 120; Gaps 15;

```

QY 95 ILMMFVGIAPALFGFGNGYLMPLQIGAPDMAFPRMNNLSFWL-----FLAGTAMGVAS 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 VLLVHGVTLPAGSNGINAVL-----HIDFVRLKEATWIDAATQIFPSLGAAGFVLI 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 LPAFGG-----DGL-----GSGVG 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 AFASYNKEDNNCYRDALLTSSINCITSPVSGFAISILGYMAHEKVNIEDVATEGAGLV 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 WVLVPLSTREAGYSMDLAFVHL--SGASSIMGAINMTTFLLMRAFGMTLHKVPLFS 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 FILPEAISTLSGSTFWAVFFVMLALGLDSSMGMEAVITGLADDFQVLKRHR-KLFT 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 WSIFITAMILLALPYLAGAITMLTDRNFGTTFENPAGGDPILYOHIL-----WFG 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 FGVTFSTFLALFCITKGIYVLTLD---TF---AAGTSILFAVLMELAGVSMFYG 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 -----HPEVY-----IILPGFIISHVSTFSKRPV--GYLPMVYMAVA 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 VDRFSNDIQOMMGRFRLGLWRLCKRFVSPALFLFVVVVSIIKFKPLTYDDYI----- 549
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 IGVLFVVAHHMYTVGMSLTQOSYFMLATWVIAVPTGIKIFSWIATMW 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 -----FPPMAN---WVGWGIALLSSMWLVPIYI-----YKFLSTQGSIM 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9
 US-08-240-783B-3
 ; Sequence 3, Application US/08240783B
 ; Patent No. 5756348

GENERAL INFORMATION:
 APPLICANT: Smith, Kelli
 APPLICANT: Borden, Laurence A.
 APPLICANT: Branchek, Theresa
 APPLICANT: Hartig, Paul R.
 APPLICANT: Weinschenk, Richard L.
 TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/240,783B
 FILING DATE:
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 617 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: N

FRAGMENT TYPE: Internal
 ORIGINAL SOURCE:
 ORGANISM: HUMAN NORADRENALINE TRANSPORTER
 US-08-240-783B-3

Query Match 3.5%; Score 106; DB 1; Length 617;
 Best Local Similarity 19.8%; Pred. No. 0.038;
 Matches 69; Conservative 43; Mismatches 117; Indels 120; Gaps 15;

```

QY 95 ILMMFVGIAPALFGFGNGYLMPLQIGAPDMAFPRMNNLSFWL-----FLAGTAMGVAS 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 276 VLLVHGVTLPAGSNGINAVL-----HIDFVRLKEATWIDAATQIFPSLGAAGFVLI 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 LPAFGG-----DGL-----GSGVG 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 AFASYNKEDNNCYRDALLTSSINCITSPVSGFAISILGYMAHEKVNIEDVATEGAGLV 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 WVLVPLSTREAGYSMDLAFVHL--SGASSIMGAINMTTFLLMRAFGMTLHKVPLFS 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 FILPEAISTLSGSTFWAVFFVMLALGLDSSMGMEAVITGLADDFQVLKRHR-KLFT 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 WSIFITAMILLALPYLAGAITMLTDRNFGTTFENPAGGDPILYOHIL-----WFG 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 447 FGVTFSTFLALFCITKGIYVLTLD---TF---AAGTSILFAVLMELAGVSMFYG 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 -----HPEVY-----IILPGFIISHVSTFSKRPV--GYLPMVYMAVA 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 VDRFSNDIQOMMGRFRLGLWRLCKRFVSPALFLFVVVVSIIKFKPLTYDDYI----- 549
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 IGVLFVVAHHMYTVGMSLTQOSYFMLATWVIAVPTGIKIFSWIATMW 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 -----FPPMAN---WVGWGIALLSSMWLVPIYI-----YKFLSTQGSIM 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10
 US-08-753-985-11

; Sequence 11, Application US/08753985
 ; Patent No. 5759788

GENERAL INFORMATION:
 APPLICANT: Fremieu Jr., Robert T.
 APPLICANT: Caron, Marc G.
 APPLICANT: Blakely, Randy D.
 TITLE OF INVENTION: A High Affinity L-Proline Transporter
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Kilpatrick & Cody
 STREET: 1100 Peachtree Street, Suite 2800
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: U.S.
 ZIP: 30309
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/753,985
 FILING DATE: 03-DEC-1996
 CLASSIFICATION: 536
 PRIOR APPLICATION NUMBER:
 APPLICATION NUMBER: US 07/879617
 FILING DATE: 01-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Pabst, Patrice L.
 REGISTRATION NUMBER: 31,284
 REFERENCE/DOCKET NUMBER: EMU109
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404-815-6508
 TELEFAX: 404-815-6555
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:

```

; LENGTH: 617 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Brain -norepinephrine transporter
;
; FEATURE:
; NAME/KEY: Active-site
; LOCATION: 95..96
; OTHER INFORMATION: /note="Leucine zipper motif"
;
; FEATURE:
; NAME/KEY: Active-site
; LOCATION: 102..103
; OTHER INFORMATION: /note="Leucine zipper motif"
;
; FEATURE:
; NAME/KEY: Active-site
; LOCATION: 109..110
; OTHER INFORMATION: /note="Leucine zipper motif"
;
; FEATURE:
; NAME/KEY: Active-site
; LOCATION: 116..117
; OTHER INFORMATION: /note="Leucine zipper motif"
;
; US-08-753-985-11

```

```

Query Match 3.5%; Score 106; DB 1; Length 617;
Best Local Similarity 19.8%; Pred. No. 0.038;
Matches 69; Conservative 43; Mismatches 117; Indels 120; Gaps 15;

```

```

QY 95 ILIMFVGIPLFGFGNYLMPLOIGAPDMAPRRNNLSFWL-----FIAGTANGVAS 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 276 VLVHGVTLPGASNGINAYL-----HIDFYRLKEATVMDAATQIFESLGAGFGVLI 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 LFAPGG-----DQGL-----GSGVG 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 AFASYNKFDNNCYRDALLTSSINCITSFVSGFATFSLIGYMAHEKVNIEDVATGAGLV 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 WVLVPLSTREAGYSMDLAIFAVHL--SGASSIMGAIMNITTFINMRAPGMLHKVPLFS 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 388 FILYPEAISTLSGSTFMVAVFVFMALLGLDSSMGMEAVITGLADDPQVLRHR-KLFT 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 WSIFITAMLLILALPYLAGATMLLTDNRNGETTFENPAGGDPILYQHIL-----WFG 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 447 FGVTFFSLALFCITKGIYVLTLD-----TF-----AAGTSILFAVLMEAGVSWFYG 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 -----HPEVY-----IILPGFGIISHVSTFSKKPVF--GYLPMVYAMA 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 498 VDRFSNDIQOMGFRPGLIYWLCKKFPVSPALLPVVYVSIINKEPLTYDDYI----- 549
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 IGVIGFVVAHMYTVGMSLTQOSYFMLATMVAVPTGIKIFSMTATMW 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 550 -----FPPMAN---WVGWGIALLSMVLVPIYI-----YKFLSTQSGSLW 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-09-084-813-3
; Sequence 3, Application US/09084813
; Patent No. 6127131
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Branchek, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weishank, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,813
; FILING DATE: 26-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-AZ-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 617 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: N
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: HUMAN NORADRENALINE TRANSPORTER
;
; US-09-084-813-3

```

```

Query Match 3.5%; Score 106; DB 3; Length 617;
Best Local Similarity 19.8%; Pred. No. 0.038;
Matches 69; Conservative 43; Mismatches 117; Indels 120; Gaps 15;

```

```

QY 95 ILIMFVGIPLFGFGNYLMPLOIGAPDMAPRRNNLSFWL-----FIAGTANGVAS 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 276 VLVHGVTLPGASNGINAYL-----HIDFYRLKEATVMDAATQIFESLGAGFGVLI 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 LFAPGG-----DQGL-----GSGVG 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 328 AFASYNKFDNNCYRDALLTSSINCITSFVSGFATFSLIGYMAHEKVNIEDVATGAGLV 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 WVLVPLSTREAGYSMDLAIFAVHL--SGASSIMGAIMNITTFINMRAPGMLHKVPLFS 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 388 FILYPEAISTLSGSTFMVAVFVFMALLGLDSSMGMEAVITGLADDPQVLRHR-KLFT 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 221 WSIFITAMLLILALPYLAGATMLLTDNRNGETTFENPAGGDPILYQHIL-----WFG 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 447 FGVTFFSLALFCITKGIYVLTLD-----TF-----AAGTSILFAVLMEAGVSWFYG 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 -----HPEVY-----IILPGFGIISHVSTFSKKPVF--GYLPMVYAMA 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 498 VDRFSNDIQOMGFRPGLIYWLCKKFPVSPALLPVVYVSIINKEPLTYDDYI----- 549
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 IGVIGFVVAHMYTVGMSLTQOSYFMLATMVAVPTGIKIFSMTATMW 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 550 -----FPPMAN---WVGWGIALLSMVLVPIYI-----YKFLSTQSGSLW 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
PCT-US92-09662-3
; Sequence 3, Application PC/TUS9209662
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Branchek, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weishank, Richard L.

```

```

: TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES THEREOF
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/09662
: FILING DATE: 19921112
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 977-9809
: TELEX: 422523 COOP UI
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 617 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEetical: N
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: ORGANISM: HUMAN NORADRENALINE TRANSPORTER
: PCT-US92-09662-3

```

Query Match 3.5%; Score 106; DB 5; Length 617;

Best Local Similarity 19.8%; Pred. No. 0.038; Mismatches 117; Indels 120; Gaps 15;

```

: 95 ILMFVGVIPALGFGFNYLPLQIGAPDMAPFRMNNLSFVL-----FIAGTAMGVAS 147
: 276 VLVHGTLPGLGASNGINAYL-----HIDFRLKEATVWIDAAQIFESLGAGFGLI 327
: 148 LRPBG-----DGL-----GSGVG 162
: 328 AFASYNKFDNNCYRDALLTSSINCITSFVSGAIFSLGYMAHEHKVNIEDVATGAGLV 387
: 163 WVIYPLSTREAGYSMDLAFVAHL--SGASSIMGAINMTTFLLNRRAPGMTLHKVPLFS 220
: 388 FIIPEAISTLSGSTWAVVFFVMLALGLDSSMGMEAVITGLADFOVLKRR- KLPF 446
: 221 WSIFITAMLLIALPVLAGATMLLTDNRNGTFFNPAGGDDILYQHIL-----WFEF 274
: 447 FGVTESFFLALFCITKGIYVLTLD-----TF-----AAGTSLPVLMEATIGVMEFG 497
: 275 -----HPEVY-----IIILPFGIISHVSTFSKRPV--GYLPMVYAMA 313
: 498 VDRFSNDIQMGFRGLWRLCWFVSPAFILFVVVVSIIINKPLTYDYI----- 549
: 314 IGVLFVNMHMYTGMSLTQSYFMLATMVLAFTGIKIFSMITMM 362
: 550 -----FPPMAN---WVGWGIALLSSMWLVPIYV-----YKFLSTQGSIM 585

```

RESULT 13

```

: US-08-191-337-3
: Sequence 3, Application US/08191337
: Patent No. 5643745
: GENERAL INFORMATION:

```

```

: APPLICANT: STUART, W. DORSEY
: TITLE OF INVENTION: HETEROLOGOUS DIMERIC PROTEINS PRODUCED
: TITLE OF INVENTION: IN HETEROOKARYONS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Ave. N.W.
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20006-1812
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/191,337
: FILING DATE: 03-FEB-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURASHIGE, KATE H.
: REGISTRATION NUMBER: 29,959
: REFERENCE/DOCKET NUMBER: 3918-0002.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
: TELEX: 90-4030
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 451 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-191-337-3

```

Query Match 3.4%; Score 103; DB 1; Length 451;

Best Local Similarity 20.0%; Pred. No. 0.049; Mismatches 154; Indels 154; Gaps 23;

```

: 109 GFGNLMPLQIGAPD--MAFPRM--NNISFWLFIAGTAMGVASLFAFGGQGLSSGVW 164
: 31 GIGGIMDEKPEAOAGNAKFRHGLKRLTVLIVAEALG--SLSPAEATYLDMPGVI 88
: 165 IYPLSTREAGYSMDLAFVAHLGASSIMGAINMTTFLLNRRAPGMT---LHKVPLFS 220
: 89 L-----SVGGL-ICITYAHVIGT-----KLKPELAHYADLRVMEGR 127
: 221 WSIFITAMLLIAL-----PVLAGATM-LTDRNGTFFNPAGGDDILYQHILMF 273
: 128 WGEYETISFMYLQILFIVGSHVLTGIMGTITDNGTCSL----- 169
: 274 GHPEVYIIILGFGIISVV-----STFSKRPVEGYLPWV-----YAMVAGV--- 316
: 170 -----VEGIVSAIILFLAIPPSFAVVALIGYIDVSCAALLIMITGRSS 218
: 317 -----LGEVVAHNMHYTGMSLTQSYFMLATMVLAFTGIKIFSMIATMGSSVERKSP 371
: 219 HQEGGLAIVPW-----SCMPREDLSLAEGLFVANSIVFAVSFAMCQSFMDENHTP 269
: 372 -----MLNAFGE--MFLTVGSGVGIYLAQGLDRAVHDTYVVAHHRVYVSLGATFAIF 424
: 270 SDYKRSIVALLGLIEIFITYTG--GVVYAFVGP-----VQSPALISAGPLAKV 317
: 425 AGIYFPMKFSGRAPENMAAKLHFWTFIGANV--TFEPQHLFGROGMPR--RYIDYPEA 480
: 318 A-----FGIALP-----VIFISGINTVVVSRILIERIMPNNVIRIVNPAG 359
: 481 FALW-----NKVSYGAFLAFASEFLF 501
: 360 WVMVGLPFGITLLIAMVIAEAIFFSDDLALICSALEF 395

```


Db 268 AFEEVGPPIPMFIVAEIIFSGPRPAIAVAGEFN-WTSNFIWMCFOYVEOLGPPYFI 326
Qy 423 IFAGI-----YFYMPKFSGRAPEMAA 444
|| : || : || : || :
Db 327 IFTVLLVLFIFTYFKVPETKGRTFDEIAS 356

Search completed: December 22, 2001, 11:02:58
Job time: 337 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 10:55:21 ; Search time 24.27 Seconds

(without alignments)
1748.217 Million cell updates/sec

Title: US-09-712-768-2

Perfect score: 3000

Sequence: 1 MADAAIGHDHHEKGGFFTR.....PAHTEETLPKRSDDMKHPSH 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2538.5	84.6	558	2	S08270 cytochrome-c oxida
2	2453.5	81.8	565	2	S20534 cytochrome-c oxida
3	2347	78.2	554	2	S03809 cytochrome-c oxida
4	1873.5	62.5	538	2	S36424 cytochrome-c oxida
5	1829	61.0	541	1	ODZ01 cytochrome-c oxida
6	1773	59.1	522	2	S62706 cytochrome-c oxida
7	1765.5	58.5	523	2	S25956 cytochrome-c oxida
8	1748.5	58.3	524	2	S14138 cytochrome-c oxida
9	1745	58.2	515	2	T11913 cytochrome-c oxida
10	1744.5	58.1	527	2	S65366 cytochrome-c oxida
11	1744	58.1	534	2	D71698 cytochrome-c oxida
12	1738.5	58.0	527	2	S14139 cytochrome-c oxida
13	1730.5	57.7	524	1	ODR21 cytochrome-c oxida
14	1730.5	57.7	524	1	ODZM1 cytochrome-c oxida
15	1729.5	57.6	524	2	S16256 cytochrome-c oxida
16	1713.5	57.1	514	2	S30281 cytochrome-c oxida
17	1707	56.9	537	2	T11237 cytochrome-c oxida
18	1698.5	56.6	531	2	S78190 cytochrome-c oxida
19	1677.5	55.9	532	2	S59087 cytochrome-c oxida
20	1677	55.9	536	2	S63651 cytochrome-c oxida
21	1670	55.7	533	2	A58931 cytochrome-c oxida
22	1651.5	55.0	527	1	OBSTY1 cytochrome-c oxida
23	1643.5	54.8	527	2	S05290 cytochrome-c oxida
24	1642	54.7	531	2	T12406 cytochrome-c oxida
25	1639	54.6	530	2	T11884 cytochrome-c oxida
26	1609	53.6	557	1	ODNC1 cytochrome-c oxida
27	1608.5	53.6	527	1	ODOB1M cytochrome-c oxida
28	1605.5	53.4	567	1	ODAS1 cytochrome-c oxida
29	1600.5	53.4	471	2	T43827 cytochrome-c oxida

30	1600	53.3	528	2	S26948 cytochrome-c oxida
31	1594	53.1	526	2	S62763 cytochrome-c oxida
32	1594	53.1	541	2	A48327 cytochrome-c oxida
33	1588.5	53.0	517	2	C34284 cytochrome-c oxida
34	1564.5	52.1	517	2	S01501 cytochrome-c oxida
35	1562.5	52.1	517	2	S14205 cytochrome-c oxida
36	1561.5	52.0	873	1	S53828 cytochrome-c oxida
37	1561	52.0	519	2	T11129 cytochrome-c oxida
38	1558.5	52.0	517	2	S70597 cytochrome-c oxida
39	1558	51.9	516	2	T11182 cytochrome-c oxida
40	1554.5	51.8	512	1	ODR11 cytochrome-c oxida
41	1552	51.7	516	2	T11195 cytochrome-c oxida
42	1550	51.7	513	2	T11482 cytochrome-c oxida
43	1548.5	51.6	534	2	S78640 cytochrome-c oxida
44	1548	51.5	516	2	S45351 cytochrome-c oxida
45	1545.5	51.5	534	2	JU0146 cytochrome-c oxida

ALIGNMENTS

RESULT 1

S08270 cytochrome-c oxidase (EC 1.9.3.1) chain I-beta - Paracoccus denitrificans

C:Species: Paracoccus denitrificans

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-May-2000

C:Accession: S08270; C35121

R:Raitto, M.; Pispas, J.M.; Metsu, T.; Saraste, M.

FEBS Lett. 261, 431-435, 1990

A:Title: Are there isoenzymes of cytochrome c oxidase in Paracoccus denitrificans?

A:Reference number: S08269; M01D:90184495

A:Accession: S08270

A:Molecule type: DNA

A:Residues: 1-558 <RAT>

A:Cross-references: EMBL:Y07533; NID:g45477; PIDN:CAA68821.1; PID:g45479

R:Van Spanning, R.J.M.; Mansell, C.; Harms, N.; Oltmann, L.F.; Stoutamer, A.H.

J. Bacteriol. 172, 986-996, 1990

A:Title: Mutagenesis of the gene encoding cytochrome c-550 of Paracoccus denitrifican

A:Reference number: A35121; M01D:90130336

A:Accession: C35121

A:Molecule type: DNA

A:Residues: 1-58 <VAN>

A:Cross-references: GB:M27304; NID:g150573; PIDN:AA88365.1; PID:g150575

R:Busse, G.; Soullame, T.; Dewor, M.; Meyer, H.E.; Bluggel, M.

Protein Sci. 8, 985-990, 1999

A:Title: Evidence for a copper-coordinated histidine-tyrosine cross-link in the activ

A:Reference number: A58960; M01D:99268331

A:Contents: annotation

A:Note: mass spectrographic and chemical characterization of histidyl-tyrosine cross-

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

F:Keywords: Chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane

F:27-492/Domain: cytochrome-c oxidase chain I homology <TM01>

F:27-59/Domain: transmembrane #status experimental <TM02>

F:84-121/Domain: transmembrane #status experimental <TM03>

F:130-151/Domain: transmembrane #status experimental <TM04>

F:178-206/Domain: transmembrane #status experimental <TM05>

F:218-251/Domain: transmembrane #status experimental <TM06>

F:263-298/Domain: transmembrane #status experimental <TM07>

F:304-322/Domain: transmembrane #status experimental <TM08>

F:334-362/Domain: transmembrane #status experimental <TM09>

F:370-395/Domain: transmembrane #status experimental <TM10>

F:404-430/Domain: transmembrane #status experimental <TM11>

F:441-468/Domain: transmembrane #status experimental <TM12>

F:483-513/Domain: transmembrane #status experimental <TM13>

F:56-80/Disulfide bonds: #status experimental

F:94,413/Binding site: heme a iron (His) (axial ligands) #status experimental

F:276,325,326/Binding site: copper (His) #status experimental

F:276-280/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status experimental

F:280/Binding site: oxygen (Tyr) #status predicted

F:403/Binding site: magnesium (His) (shared with chain II) #status predicted

F:411/Binding site: heme a3 iron (His) (axial ligand) #status experimental

```

Query Match      84.6%  Score 2538.5; DB 2; Length 558;
Best Local Similarity 81.0%; Pred. No. 5e-169;
Matches 452; Conservative 51; Mismatches 54; Indels 1; Gaps 1;

Oy 1 MADAAIHGH-DHNEKOGFTRMFMSTNKHDKGLLYLVAGVVGFTSYLFYVMRLMLDP 59
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MADAAVHGHGHHDRKGFTRFMSTNKHDKGLLYLVAGVVGFTSYLFYVMRLMLDP 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 60 GVQYMCLEGARLIADASOTCTANGHLMNVVTVYHGIIMLFEPVGIIPALFEGFGNYLPLQI 119
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 GVQYMCLEGARLIADASOCTPRNGHLMNVVTVYHGIIMLFEPVGIIPALFEGFGNYLPL 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 120 GAPDMAFRRMNNLSWFLFAGTAMGVASLFRPGDGOLGSGVGNVLPRLSTRRAGSMD 179
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GAPDMAFRRLNLSWVWVAGVAVASLFRPGDNDGSGVGNVLPRLSTRRAGSMD 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 180 LAIFVNHSGASSINGAINMTTFPLNMPAPGNTLHKVPLFESMSIFITPMLLLALPYLAG 239
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 LAIFVNHSGASSILGAINIITTFPLNMPAPGNTLHKVPLFESMSVFTTMLLLSLPYLAG 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 240 AITMLTRNRFETFPFNAGGDDPILYOHILMFEGHREPVYIIILPFGIISHVSTESK 299
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 AITMLMRNFGTOQFPRDAGGDDPYLYOHILMFEGHREPVYIIILPFGIISHVISTFAKK 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 300 PYEGCLPVMYANVAIGVLCFVYVVAHNMHTVGMSLTQDSYFMLATNVIAVPRGKIFESIA 359
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 PYEGCLPVMYLAAGVIGVLCFVYVVAHNMHTVGMSLTQDSYFMLATNVIAVPRGKIFESIA 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 360 TWMGSSVEFKSPMLAFGFMFLFTVGVTGVLIAQAGIDRAYNHDYVVAHFHVMISGA 419
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 TWMGSSIEFKTMLAFGFLFETYGVTGVVLSAPRLDRKYHHDYVVAHFHVMISGA 420
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 420 IFAIFAGIETVMPKFSGNAFREPMAAKLHFWTFEGANVTEFPQHELGQGNPRRYIDPE 479
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 VEGIFAGVYVWIKKSGHQYREMAQLHFMMWIFGNSNLIFPPQHLRGQMPRRYIDPV 480
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 480 AFALNNKVSYGAFPLAFASPLFETIYVTVLAGRRETRPMPGFAATLEWTLPSPPA 539
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 EFAYNNNLSIGAYISFASPLFETIYVTVLAGRKNVAVPNVMEHADTLEWTLPSPE 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 540 HFEETLPKRSDMDKHPSH 557
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 HFEETLPKRSDMDRAAH 558
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
S20534
cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhodobacter sphaeroides
C:Species: Rhodobacter sphaeroides
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text-change 20-Aug-1999
C:Accession: S20534; A42378
R:Shapleigh, J. P.; Gennis, R. B.
MOL. Microbiol. 6, 635-642, 1992
A>Title: Cloning, sequencing and deletion from the chromosome of the gene encoding subunit
A:Reference number: S20534; MUID:92204019
A:Accession: S20534
A:Molecule type: DNA
A:Residues: 1-565 <SHA>
A:Cross-references: EMBL:X62645; NID:946417; PID:CAA4514.1; PID:946418
A>Note: the authors translated the codon AAC for residue 214 as Met and ATG for residue
R:Shapleigh, J. P.; Hill, J. V.; Alden, J. O.; Gennis, R. B.
J. Bacteriol. 174, 2338-2343, 1992
A>Title: Spectroscopic and genetic evidence for two heme-Cu-containing oxidases in Rhodospirillum rubrum
A:Reference number: A42378; MUID:92202164
A:Accession: A42378
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 286-289, 'LT', 292-293, 'G', 295-296, 'MT', 299, 'QIVS', 304-305, 'SR', 308-309, 'V', 310-311, 'AARLN', 'S', 486, 'S', 488, 'S' <SH2>
C:Genetics:
A:Gene: ctad
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

```

C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane
F:25-500/Domain: cytochrome-c oxidase chain I homology <COI>
F:102,421/Binding site: heme a iron (His) (axial ligands) #status predicted
F:284,333,334/Binding site: copper (His) #status predicted
F:284-288/Cross-link: 1'-histidyl-3'-lysine (His-Tyr) #status predicted
F:288/Binding site: oxygen (Tyr) #status predicted
F:419/Binding site: magnesium (His) (shared with chain II) #status predicted
F:419/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match . 81.8%; Score 2453; DB 2; Length 565;
Best Local Similarity 78.7%; Pred. No. 4,4e-163;
Matches 446; Conservative 50; Mismatches 59; Indels 12; Gaps 3;

```

QY      1  MADAALHGHDDHEKOGFTTRMFSTNKKDGLLTLVAAGVGFISVLFYMYRLMDPG 60
DB      1  MADAALHGHHE-HDRGRFTFRMFSTNKKDGLVLFPGGGLGLLSVAFTMYRMELIAP 59
QY      61  VOYMLLEGAR-----LIADASOTCTANGHLMNMYVYHGLIMLMEFGIALRGGF 110
DB      60  VOYMLAEHLESLVAGFGFQSLMPSAVENTCTRGHLMNMYITGHILIMLFFVYIPALRGGF 119
QY      111 GNYLMPLOIGADMAFPRMNNLSFWLFTAGTAMGVASLFARGGDGLSGSGVWLYPLLS 170
DB      120 GNYFMPRLHIGADMAFPRMNNLSYVLVAGTSLVASLFAFGNGDGLSGGWLVLPLLS 179
QY      171 TREAGYSMDLAFVAHLSGASSIMGATIMITFTPLNMRAPGWTLHKVLFMSITFTANLI 230
DB      180 TSEGSTDLAFVAHLSGASSILGATIMITFTPLNMRAPGWTMHRKPLFASITFTANLI 239
QY      231 LLAEVLGATIMTLTDENFGTTFPNPAGGDDPLLYOHLIMLFGHPEVYIILRFGIIS 290
DB      240 LLALPVLGATIMTLTDENFGTTFPGSGGDDPLYLCHILMEFGHPEVYIILPAFGIS 299
QY      291 HVSFTFSKKRPVPGYLPNYTAMVAGLVGFVYVAHNMITYGSLTQOSTFMLATGVIAPT 350
DB      300 HVIATFPAKRPLEGYLPNYTAMVAGLVGFVYVAHNMITYAGSLTQOSTFMAITGVIAPT 359
QY      351 GIKTFSMATIMMGSGVEKSPMLMAFGFMLEFVYGVGTVGLAQAGLDRAVNDHYVVAH 410
DB      360 GIKTFSMATIMMGSSIELKTPMLALGFLFLFYGVGTVGLVLSQASDRTYNDHYVVAH 419
QY      411 FHYVMSLGAIEAIEAGVIEYMPKESGRAFPKMAKLIHFTEFGANVTFFPHFLGRQM 470
DB      420 FHYVMSLGAIEGIEIFAGTSGIKMGSRQYRPMAGLIFHMFMEVAVANLFFPHFLGRQM 479
QY      471 PRRYIDYPAAPFLMKVSSYGAFLAFASEFLFEIYFVYTLVAGRETRPNPMSGFADTLE 530
DB      480 PRRYIDYPAAPFLMVFVSSLGAFLSFASEFLFEIYFVSL-SGARVTANNMYNEHADTLE 538
QY      531 WTLPSPPAHNPFETLPKRSDMDKHNPS 557
DB      539 WTLSPRPENHFEQLPKREDMERAPAH 565

```

RESULT 3
S03809
cytochrome-c oxidase (EC 1.9.3.1) chain I - Paracoccus denitrificans
C:Species: Paracoccus denitrificans
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Jun-2000
C:Accession: S03809
R:Raitio, M.; Jalli, T.; Saraste, M.
EMBO J. 6, 2825-2833, 1987
A>Title: Isolation and analysis of the genes for cytochrome c oxidase in Paracoccus d
A:Reference number: S03803
A:Accession: S03809
A:Molecule type: DNA
A:Residues: 1-554 <RAI>
A:Cross-references: EMBL:X05829; NID:9994806; PIDN:CAA29274.1; PID:9995678
C:Genetics:
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane

F:24-487/Domain: cytochrome-c oxidase chain I homology <CO1>
 F:91,408/Binding site: heme a iron (His) (axial ligand) #status predicted
 F:273,322,333/Binding site: copper (His) #status predicted
 F:273-277/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:277/Binding site: oxygen (Tyr) #status predicted
 F:398/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:406/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 78.2% Score 2347; DB 2; Length 554;

Best Local Similarity 76.1% Pred. No. 9.9e-156; Matches 415; Conservative 59; Mismatches 69; Indels 2; Gaps 1;

QY 13 EKQGFETRMSTNKHDKGLYLVAAGVGFISVLFVYMRLELMDPQVYMCLEGARLI 72
 DB 11 EKRGFTFMSTNKHDKGLYLVAAGVGFISVLFVYMRLELMDPQVYMCLEGARLI 70
 QY 73 ADASOTCTANGHLNWNVYTHGILMFFVGPALFSGFNYLMPLOIGAPDAPRRNNL 132
 DB 71 ADAAECTPRNMLNWNVYTHGILMFFVGPALFSGFNYLMPLOIGAPDAPRRNNL 130
 QY 133 SFWLEIAGTANGVSLFAPGDDGOLGSGGVNLYPLSTRAGYSMDLAIFAVHLSGASS 192
 DB 131 SYWLVCVSLAISLSLSPGSDOPGAGVNLPLSTRAGYAMDLAIFAVHLSGATS 190
 QY 193 IMAINMTTFLNMRAPGMLTHKVPLEFSMIFITAMILLALPVLACATMLLTDNRFGT 252
 DB 191 ILGAINITTFELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLACATMLLTDNRFGT 250
 QY 253 TFFNAGGDDPLVOHILMFFGHPREYIILPGGITSHVVSFESKRPVEGYLPMYAMV 312
 DB 251 OFEDPAGGDDPLVOHILMFFGHPREYIILPGGITSHVVSFESKRPVEGYLPMYAMA 310
 QY 313 AIGVGFVMAHMYTVMSLTQOSTFMLATMVAVPTGIFESFIATMWSGVEFSKPM 372
 DB 311 AIAFGFTVMAHMYTVMSLTQOSTFMLATMVAVPTGIFESFIATMWSGVEFSKPM 370
 QY 373 LMAFGFMFLTVGGVTVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAFAAGIYFMP 432
 DB 371 LMAFLA-FLFTVGGVTVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAFAAGIYVIG 428
 QY 433 KESGAFPEBMAAKLHFWTFEFGANVTFFPHFLRGQMPRRYIDYDEAFALMKNVSSYGA 492
 DB 429 KMSGQVPEBMAAKLHFWTFEFGANVTFFPHFLRGQMPRRYIDYDEAFALMKNVSSYGA 488
 QY 493 FLAFASTFLFYIYVTVLAGRRETRPNMGEPADLEMTLPSPPAATFETLKRSDMD 552
 DB 489 YISFASSTFLFYIYVTVLAGRRETRPNMGEPADLEMTLPSPPAATFETLKRSDMD 548
 QY 553 KHPSH 557
 DB 549 RQAQH 553

RESULT 4
 S36424
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhizobium leguminosarum

C:Species: Rhizobium leguminosarum
 C:Date: 09-Dec-1993 #sequence_revision 03-Nov-1995 #text_change 23-Jul-1999
 C:Accession: S36424
 R:Gabel, C.; Bittlinger, M.A.; Maier, R.J.
 submitted to the EMBL Data Library, July 1993
 A:Description: Cytochrome aa3 gene regulation in the Rhizobiaceae: comparison of copper
 A:Reference number: S36424
 A:Accession: S36424
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-538 <GAB>
 A:Cross-references: EMBL:X74341; NID:g396698; PIDN:CAA52388.1; PID:g396699
 C:Genetics:
 A:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as

rotein
 F:31-481/Domain: cytochrome-c oxidase chain I homology <CO1>
 F:82,402/Binding site: heme a iron (His) (axial ligand) #status predicted
 F:265,314,315/Binding site: copper (His) #status predicted
 F:265-269/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:269/Binding site: oxygen (Tyr) #status predicted
 F:392/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:400/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 62.5% Score 1873.5; DB 2; Length 538;

Best Local Similarity 64.2% Pred. No. 7.2e-123; Matches 348; Conservative 63; Mismatches 110; Indels 21; Gaps 4;

QY 7 HGHDHHEKQGFETRMSTNKHDKGLYLVAAGVGFISVLFVYMRLELMDPQVYMCLE 66
 DB 15 HGHDEHAP--TGWRSTNKHDKGLYLVAAGVGFISVLFVYMRLELMDPQVYMCLE 69
 QY 67 EGARLIADASOTCTANGHLNWNVYTHGILMFFVGPALFSGFNYLMPLOIGAPDAP 126
 DB 70 -----HNHILNWNVYTHGILMFFVGPALFSGFNYLMPLOIGAPDAP 115
 QY 127 PRANNISFWLPIAGTANGVSLAPGDDGOLGSGGVNLYPLSTR-EGAGSMDLAIFAV 185
 DB 116 PRANNISFWLPIAGTANGVSLAPGDDGOLGSGGVNLYPLSTR-EGAGSMDLAIFAV 175
 QY 186 HLGASSIMCAINMTTFELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLACATML 245
 DB 176 HLGASSILGAINITTFELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLACATML 235
 QY 246 TDRNFGTFFNAGGDDPLVOHILMFFGHPREYIILPGGITSHVVSFESKRPVEGYL 305
 DB 236 TDRNFGTFFNAGGDDPLVOHILMFFGHPREYIILPGGITSHVVSFESKRPVEGYL 295
 QY 306 RMYAMAVIGLVGVVAHMYTVMSLTQOSTFMLATMVAVPTGIFESFIATMWSG 365
 DB 296 RMYAMAVIGLVGVVAHMYTVMSLTQOSTFMLATMVAVPTGIFESFIATMWSG 355
 QY 366 VEPKSPMLAFGEFLETVGGVTVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAIFA 425
 DB 356 IEPATPMMLAFLEFLETVGGVTVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAIFA 415
 QY 426 GIFYMPEKSGRAPEBMAAKLHFWTFEFGANVTFFPHFLRGQMPRRYIDYDEAFALM 485
 DB 416 GMYWPEKSGRAPEBMAAKLHFWTFEFGANVTFFPHFLRGQMPRRYIDYDEAFALM 475
 QY 486 KVSYSYGAFLAFSTFLFYIYVTVLAGRRETRPNMGEPADLEMTLPSPPAATFETL 545
 DB 476 LVSSISGYSISGFSVLYIC-VYDAFAKNVPVGDNPAGATLLEMTLPSPPVHEFEVL 534
 QY 546 PK 547
 DB 535 PR 536

RESULT 5
 ODZJ1
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Bradyrhizobium japonicum

C:Species: Bradyrhizobium japonicum
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 23-Jul-1999
 C:Accession: S13076; S12101
 R:Botl, M.; Bolliger, M.; Hennecke, H.
 Mol. Microbiol. 4, 2147-2157, 1990
 A:Title: Genetic analysis of the cytochrome c-aa(3) branch of the Bradyrhizobium japo
 A:Reference number: S13076; M01D:91211625
 A:Accession: S13076
 A:Molecule type: DNA
 A:Residues: 1-541 <BOT>
 A:Cross-references: EMBL:X54800; NID:g39502; PIDN:CAA38570.1; PID:g39503
 R:Gabel, C.; Maier, R.J.
 Nucleic Acids Res. 18, 6143, 1990
 A:Title: Nucleotide sequence of the cox a gene encoding subunit I of cytochrome aa(3)
 A:Reference number: S12101; M01D:91045095

A:Accession: S12101
 A:Molecule type: DNA
 A:Residues: 1-541 <CAB>
 A:Cross-references: EMBL:X54318; NID:g39505; PIDN:CAA3821.1; PID:g39506
 C:Genetics:
 A:Gene: coxa
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
 rotein
 F:34-484/Domain: cytochrome-c oxidase chain I homology <CO1>
 F:41-60/Domain: transmembrane #status predicted <TM01>
 F:67-107/Domain: transmembrane #status predicted <TM02>
 F:124-142/Domain: transmembrane #status predicted <TM03>
 F:173-193/Domain: transmembrane #status predicted <TM04>
 F:211-231/Domain: transmembrane #status predicted <TM05>
 F:262-282/Domain: transmembrane #status predicted <TM06>
 F:294-315/Domain: transmembrane #status predicted <TM07>
 F:331-352/Domain: transmembrane #status predicted <TM08>
 F:365-385/Domain: transmembrane #status predicted <TM09>
 F:404-424/Domain: transmembrane #status predicted <TM10>
 F:441-460/Domain: transmembrane #status predicted <TM11>
 F:483-503/Domain: transmembrane #status predicted <TM12>
 F:545-605/Binding site: heme a iron (His) (axial ligand) #status predicted
 F:568-617/Binding site: copper (His) #status predicted
 F:268-272/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:272/Binding site: oxygen (Tyr) #status predicted
 F:403/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 61.0%; Score 1829; DB 1; Length 541;
 Best Local Similarity 61.8%; Pred. No. 8,8e-120;

Matches 337; Conservative 72; Mismatches 112; Indels 24; Gaps 4;

7 HGDHHEKQGFRTWFTSTNKHDKIDGLLYAAAGVFISLVFVVRLELMDPGVQYML 66
 15 HGDENAHNPRGMRYYSTNKHDKIDGLYFAVIAVYAGVAMSAIRAEMLPGVQIF-- 72
 67 EGARLADASQSTANGHLLNNVYTHGILMFVFGIPALFGSGFVYVPLQIGADMAF 126
 73 -----HEHTYVNEVTSGLIMIFPMVMPAMIGSGFMVPLMIGADMAF 118
 127 PRNNNSLFWLFTIGTAMGVASLEAPGGDGLSGVGVWLYPRPLSTR-EGAGSMDLAIFAV 185
 119 PRNNNSLFWLFTIGTAMGVASLEAPGGDGLSGVGVWLYPRPLSTR-EGAGSMDLAIFAV 178
 186 HLGASSIMCAIMMTFTFLMRRAPGKTLHKVPLFSISITAVLLILALPVLAGATMML 245
 179 HLGASSIIGAINPITTFIMRRAPGKTLHKVPLFSISITAVLLILALPVLAGATMML 238
 246 TDRNFGTTFEPNAGGDPILYQHILMFEGHPEVYIIILPGFGLISHVSTFSKRPVGYL 305
 239 TDRNFGTTFEPNAGGDPILYQHILMFEGHPEVYIIILPGFGLISHVSTFSKRPVGYL 298
 306 PMYVAMATIGLVPMVAMHMTVGMSTLQOQSEMLATVIANPFCIKIPSWATMGGS 365
 299 GMVYAMATIGLVPMVAMHMTVGMSTLQOQSEMLATVIANPFCIKIPSWATMGGS 358
 366 VERKSPMLAMFGFMPLEFVGVGTGIVLAQAGLRAHNDTYVVAHRYVMSLGAIFA 425
 359 IERRAPIMAVGFIPLFTVGVGTGIVLAQAGLRAHNDTYVVAHRYVMSLGAIFA 418
 426 GIVYMKESGARPEWAKLHMTFTFGANVTFPQHLGROGMRRYIDPEAFALNN 485
 419 GMYVMPKMGYMNNETLAKAHMVFIFGVNLVFPQHLGSLGMRRYDYDPAFAGNN 478
 486 KVSSTYCAFLAFSLFFVIVTTLVAGRETRP---NPMGEADPLLENTLPSPPAHNF 542
 479 LVSSVSGYISG---FGVLLIFLCVIDAFKAKVPAAGDNPWAGATTLLEMTLPSPPHOF 534
 543 ETLPR 547
 535 EVLPR 539

RESULT 6

S62706

Cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) ml

C:Species: mitochondrion Platymonas subcordiformis

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text-change 23-Jul-1999

C:Accession: S62706

R:Kessler, U.; Zetsche, K.

Plant Mol. Biol. 29, 1081-1086, 1995

A:Title: Physical map and gene organization of the mitochondrial genome from the unic

A:Reference number: S62702; MUID:56145517

A:Accession: S62706

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-523 <KES>

A:Cross-references: EMBL:Z47795; NID:633584; PIDN:CAA87753.1; PID:633589

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995

C:Genetics:

A:Gene: col

A:Genome: mitochondrion

A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molec

les from the mitochondrial matrix producing two molecules of water and lowering the co

A:Pathway: oxidative phosphorylation; respiratory chain

A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnes

active phosphorylation; oxidoreductase; respiratory chain; transmembrane protein

F:1-13/Domain: mitochondrial matrix #status predicted <MM1>

F:13-460/Domain: cytochrome-c oxidase chain I homology <CO1>

F:14-42/Domain: transmembrane #status predicted <TM01>

F:54-54/Domain: intracristal #status predicted <ITC1>

F:55-90/Domain: transmembrane #status predicted <TM02>

F:91-98/Domain: mitochondrial matrix #status predicted <MM2>

F:99-121/Domain: transmembrane #status predicted <TM03>

F:122-144/Domain: intracristal #status predicted <ITC2>

F:145-174/Domain: transmembrane #status predicted <TM04>

F:175-186/Domain: mitochondrial matrix #status predicted <MM3>

F:187-216/Domain: transmembrane #status predicted <TM05>

F:217-231/Domain: intracristal #status predicted <ITC3>

F:232-265/Domain: transmembrane #status predicted <TM06>

F:266-272/Domain: mitochondrial matrix #status predicted <MM4>

F:273-288/Domain: transmembrane #status predicted <TM07>

F:299-301/Domain: intracristal #status predicted <ITC4>

F:302-330/Domain: transmembrane #status predicted <TM08>

F:331-338/Domain: mitochondrial matrix #status predicted <MM5>

F:339-359/Domain: transmembrane #status predicted <TM09>

F:360-373/Domain: intracristal #status predicted <ITC5>

F:374-403/Domain: transmembrane #status predicted <TM10>

F:404-436/Domain: mitochondrial matrix #status predicted <MM6>

F:437-449/Domain: intracristal #status predicted <ITC6>

F:450-481/Domain: transmembrane #status predicted <TM12>

F:482-522/Domain: mitochondrial matrix #status predicted <MM7>

F:523-581/Binding site: heme a iron (His) (axial ligand) #status predicted

F:582-594/Binding site: copper (His) #status predicted

F:244-248/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted

F:248/Binding site: oxygen (Tyr) #status predicted

F:322/Binding site: myristate (Lys) (covalent) #status predicted

F:371/Binding site: magnesium (His) (shared with chain II) #status predicted

F:379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 59.1%; Score 1773; DB 2; Length 523;
 Best Local Similarity 60.1%; Pred. No. 6,6e-116;

Matches 322; Conservative 82; Mismatches 104; Indels 28; Gaps 6;

17 FRTKFMSTNKHDKIDGLLYAAAGVGTISLVFTYVVRLELMDPGVQYMLGEGARLADAS 76
 4 FAQRWLFSTNKHDKIDGLYFAVIAVYAGVAMSAIRAEMLPGVQIF--N 48
 77 QCTANGHLLNNVYTHGILMFVFGIPALFGSGFVYVPLQIGADMAFPRNNLSLFWL 136

Db 49 QVLGNHQLYLVITTAHAFLEFWMVPMALLGFGNMFVPLMIGAPDMAFPRLLNNISFWL 108
 QY 137 FIAGTANGVASLFAFGDGLGSGVGVNLYPLST--REAGYSMDLAFVAHLSGASSIM 194
 Db 109 LPPSLLLLLLSALV-----EIGAGTGWTVYPPPLSSASHSGAVDLAFSLHISGASSIL 163
 QY 195 GAINMTTFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVAGATMTLLDRNFGTTF 254
 Db 164 GSNIFETITFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVAGATMTLLDRNFGTTF 223
 QY 255 FNPAGGDDPLLYOHILMEFEGHPEVYIIILPGGIIISHVSTESKRPVGYLPMVYAMAI 314
 Db 224 FDPAGGDDPLLYOHILMEFEGHPEVYIIILPGGIIISHVSTESKRPVGYLPMVYAMAI 283
 QY 315 GVLFGEVVAHMYTVGMSLTQOQSYEMLATVAVPTGIRKISWTATMGGSVSEKSPMLM 374
 Db 284 GILGFTVAHMYTVGMSLTQOQSYEMLATVAVPTGIRKISWTATMGGSVSEKSPMLM 343
 QY 375 AFGEFMEFLTVGVTGIVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAFAGIYFMPKF 434
 Db 344 AIGFLEFLTVGVTGIVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAFAGIYFMPKF 403
 QY 435 SGRAPFEMAARKLHEWTFEGICANVTFEPQHFLGROGMRRYIDYEPALMKNKVSYGAF 494
 Db 404 CGLOISELGOIHFWITFEFVNLTFEPHFLGISCMPRRIDYDPAFAGMAVSGSYL 463
 QY 495 AFASFLEFIVTVYVTVAGRRETRPNP--GEFADTLEWTLSPPAHFEFLP 546
 Db 464 SVLGALEFFVYVVDITLSGEGRCRNPWEYVPGTSA-TLEMLTSPPAHFEFLP 517

RESULT 7
 525956
 cytochrome-c oxidase (EC 1.9.3.1) chain I - liverwort (Marchantia polymorpha) mitochondrion

C:Species: mitochondrion Marchantia polymorpha
 C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 23-Jul-1999
 C:Accession: S25956
 R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kan, J. Mol. Biol. 223, 1-7, 1992
 A:Title: Gene organization deduced from the complete sequence of liverwort Marchantia p
 A:Reference number: S25941; MUID:92114051
 A:Accession: S25956
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-522 <O>A>
 A:Cross-references: EMBL:M68929; NID:9786182; PIDD:AC09451.1; PIDD:9786237
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
 C:Genetics:
 A:Gene: cox1
 A:Genome: mitochondrion
 A:Introns: 15/2; 60/1; 125/3; 132/2; 171/1; 208/3; 243/3; 372/3; 435/3
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
 transmembrane protein
 F:13-460/Domain: cytochrome-c oxidase chain I homology <COI>
 F:65,381/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:244,293,294/Binding site: copper (His) #status predicted
 F:244-248/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:248/Binding site: oxygen (Tyr) #status predicted
 F:371/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 58.5%; Score 1756.5; DB 2; Length 522;
 Best Local Similarity 58.9%; Pred. No. 9.2e-115;
 Matches 315; Conservative 92; Mismatches 101; Indels 27; Gaps 6;

QY 17 FETRFMTNKKDGLGLYLVAAGVGFISVLFVYMKLEMDPGVOYMCLEGARLIADAS 76
 Db 4 FAQRFLFSTNKKDGLGLYLVAAGVGFISVLFVYMKLEMDPGVOYMCLEGARLIADAS 48
 QY 77 QCTRANGHLMMVWYTHGILMMEFVGIPALFGGNGVLMPLDAGDPAFPRMNNLSFWL 136
 Db 51 GGNHQLYLVITTAHAFLEFWMVPMALLGFGNMFVPLMIGAPDMAFPRLLNNISFWL 108

Db 49 QVLGNHQLYLVITTAHAFLEFWMVPMALLGFGNMFVPLMIGAPDMAFPRLLNNISFWL 108
 QY 137 FIAGTANGVASLFAFGDGLGSGVGVNLYPLST--REAGYSMDLAFVAHLSGASSIM 194
 Db 109 LPPSLLLLLLSALV-----EIGAGTGWTVYPPPLSSASHSGAVDLAFSLHISGASSIL 163
 QY 195 GAINMTTFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVAGATMTLLDRNFGTTF 254
 Db 164 GSNIFETITFLNMRAPGMLTKHVPLEFSWISFTTAMLLALPLVAGATMTLLDRNFGTTF 223
 QY 255 FNPAGGDDPLLYOHILMEFEGHPEVYIIILPGGIIISHVSTESKRPVGYLPMVYAMAI 314
 Db 224 FDPAGGDDPLLYOHILMEFEGHPEVYIIILPGGIIISHVSTESKRPVGYLPMVYAMAI 283
 QY 315 GVLFGEVVAHMYTVGMSLTQOQSYEMLATVAVPTGIRKISWTATMGGSVSEKSPMLM 374
 Db 284 GILGFTVAHMYTVGMSLTQOQSYEMLATVAVPTGIRKISWTATMGGSVSEKSPMLM 343
 QY 375 AFGEFMEFLTVGVTGIVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAFAGIYFMPKF 434
 Db 344 AIGFLEFLTVGVTGIVLAQAGLDRAVHDYVVAHFHYVMSLGAIFAFAGIYFMPKF 403
 QY 435 SGRAPFEMAARKLHEWTFEGICANVTFEPQHFLGROGMRRYIDYEPALMKNKVSYGAF 494
 Db 404 CGLOISELGOIHFWITFEFVNLTFEPHFLGISCMPRRIDYDPAFAGMAVSGSYL 463
 QY 495 AFASFLEFIVTVYVTVAGRRETRPNP--GEFADTLEWTLSPPAHFEFLP 546
 Db 464 SVLGALEFFVYVVDITLSGEGRCRNPWEYVPGTSA-TLEMLTSPPAHFEFLP 516

RESULT 8
 S14138
 cytochrome-c oxidase (EC 1.9.3.1) chain I - sugar beet mitochondrion

C:Species: mitochondrion Beta vulgaris var. altissima (sugar beet)
 C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 23-Jul-1999
 C:Accession: S14138
 R:Send, M.; Harada, T.; Mikami, T.; Sugijura, M.; Kinoshita, T.
 Curr. Genet. 19, 175-181, 1991
 A:Title: Genomic organization and sequence analysis of the cytochrome oxidase subunit
 A:Reference number: S14138; MUID:91330331
 A:Accession: S14138
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-524 <SEN>
 A:Cross-references: GB:X57693; GB:S47702; NID:911258; PIDD:CAA40874.1; PIDD:911259
 C:Genetics:
 A:Gene: cox1
 A:Genome: mitochondrion
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane
 transmembrane protein
 F:12-459/Domain: cytochrome-c oxidase chain I homology <COI>
 F:64,380/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:243,292,293/Binding site: copper (His) #status predicted
 F:243-247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:247/Binding site: oxygen (Tyr) #status predicted
 F:370/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 58.3%; Score 1748.5; DB 2; Length 524;
 Best Local Similarity 59.2%; Pred. No. 3.3e-114;
 Matches 315; Conservative 86; Mismatches 106; Indels 25; Gaps 5;

QY 20 RWEFMTNKKDGLGLYLVAAGVGFISVLFVYMKLEMDPGVOYMCLEGARLIADAS 79
 Db 6 RWEFMTNKKDGLGLYLVAAGVGFISVLFVYMKLEMDPGVOYMCLEGARLIADAS 50
 QY 80 TANGCHLMNVWYTHGILMMEFVGIPALFGGNGVLMPLDAGDPAFPRMNNLSFWL 139
 Db 51 GGNHQLYLVITTAHAFLEFWMVPMALLGFGNMFVPLMIGAPDMAFPRLLNNISFWL 108

Thu Dec 27 08:21:34 2001

us-09-712-768-2.rpr

Page 10

Search completed: December 22, 2001, 11:02:25
Job time: 424 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:02:31 ; Search time 15.26 seconds
(without alignments)
1338.290 Million cell updates/sec

Title: US-09-712-768-2
Perfect score: 3000
Sequence: 1 MADDAIHGHDKHKGQFTR.....PAHTFETLPKRSMDCKHPSH 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_39.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2538.5	84.6	558	1	CX1B_PARDE
2	2453	81.8	565	1	COX1_RHOSH
3	2347	78.2	554	1	CX1A_PARDE
4	1873.5	62.5	538	1	COX1_RHILE
5	1829	61.0	541	1	COX1_BRAJA
6	1756.5	58.5	522	1	COX1_MARPO
7	1748.5	58.3	524	1	COX1_BERVU
8	1744	58.1	534	1	COX1_RICPR
9	1738.5	58.0	527	1	COX1_ARATH
10	1730.5	57.7	524	1	COX1_ORYSA
11	1730.5	57.7	528	1	COX1_ORYZA
12	1730.5	57.6	530	1	COX1_SORBI
13	1729.5	57.6	524	1	COX1_WHEAT
14	1713.5	57.1	514	1	COX1_PROMT
15	1677.5	55.9	532	1	COX1_CHOCR
16	1677	55.9	536	1	COX1_ALIMA
17	1651.5	55.0	527	1	COX1_ALOMA
18	1643.5	54.8	527	1	COX1_PEA
19	1639	54.6	530	1	COX1_SOYBN
20	1609	53.6	530	1	COX1_METSE
21	1608.5	53.6	527	1	COX1_NEUCR
22	1605.5	53.5	527	1	COX1_OENBE
23	1600	53.3	528	1	COX1_EMENT
24	1594	53.1	526	1	COX1_CVACA
25	1594	53.1	541	1	COX1_PODAN
26	1593	53.1	492	1	COX1_PHYME
27	1588.5	53.0	517	1	COX1_PARLI
28	1564.5	52.1	517	1	COX1_STRPU
29	1562.5	52.1	517	1	COX1_PISOC
30	1561.5	52.0	873	1	COX1_ACACA
31	1559	52.0	517	1	COX1_MYXGL
32	1558.5	52.0	517	1	COX1_ASTPE
33	1554.5	51.8	512	1	COX1_YEAST

34	1550	51.7	513	1	COX1_RABIT	079429 oryctolagus
35	1547	51.6	516	1	COX1_CARAU	078681 catarrhus a
36	1546	51.5	516	1	COX1_GADMO	036775 gadus morhua
37	1545.5	51.5	534	1	COX1_SACDO	P98001 saccharomyc
38	1545	51.5	516	1	COX1_STRCA	021399 struthio ca
39	1544	51.5	514	1	COX1_CERSI	003198 ceratotheri
40	1544	51.5	514	1	COX1_SHEEP	078749 ovis aries
41	1543	51.4	513	1	COX1_DIDMA	P41310 diodelphis m
42	1543	51.4	514	1	COX1_BOVIN	P00396 bos taurus
43	1543	51.4	516	1	COX1_CROLA	P34188 crossostoma
44	1542	51.4	514	1	COX1_FELCA	P48888 felis silve
45	1542	51.4	514	1	COX1_RAT	P05503 rattus norv

ALIGNMENTS

RESULT	ID	STANDARD	PRT	558 AA.
1	CX1B_PARDE			
AC	P98002;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CYTOCHROME C OXIDASE POLYPEPTIDE I-BETA (EC 1.9.3.1) (CYTOCHROME AA3 SUBUNIT 1-BETA).			
GN	CTADII.			
OS	Paracoccus denitrificans.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;			
OC	Paracoccus.			
OX	NCBI_TaxID=266;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=PD 1222;			
RX	MEDLINE=90184495; PubMed=2155830;			
RA	Raitio M., Pispas J.M., Melse T., Saraste M.;			
RT	"Are there isoenzymes of cytochrome c oxidase in Paracoccus denitrificans?";			
RT	FEBS Lett. 261:431-435(1990).			
RL	[2]			
RN	X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).			
RP	MEDLINE=95379947; PubMed=7651515;			
RX	Iwata S., Ostermeier C., Ludwig B., Michel H.;			
RA	"Structure at 2.8-A resolution of cytochrome c oxidase from Paracoccus denitrificans.";			
RT	Nature 376:660-665(1995).			
RL	[3]			
RN	X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).			
RP	MEDLINE=98021406; PubMed=9380672;			
RX	Ostermeier C., Harrenga A., Ermler U., Michel H.;			
RA	"Structure at 2.7-A resolution of the Paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody FV fragment.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 94:10547-10553(1997).			
RL	[4]			
RN	COVALENT BOND.			
RP	MEDLINE=99268331; PubMed=10338009;			
RX	Buse G., Soullimane T., Dewor M., Meyer H.E., Blueggel M.;			
RA	"Evidence for a copper-coordinated histidine-tyrosine cross-link in the active site of cytochrome oxidase.";			
RT	Protein Sci. 8:985-990(1999).			
RL	[5]			
RN	REVIEW.			
RP	MEDLINE=96391941; PubMed=8794157;			
RX	Ostermeier C., Iwata S., Michel H.;			
RA	"Cytochrome c oxidase.";			
RT	Curr. Opin. Struct. Biol. 6:460-466(1996).			
RL	[6]			
RN	FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA HEME A AND CU(A) TO THE BINUCLEAR CENTER FORMED BY HEME A3 AND CU(B). THIS CYTOCHROME C OXIDASE SHOWS PROTON PUMP ACTIVITY ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER. -I- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +			

CC 4 FERRICYTOCHROME C.
 CC COFACTOR: TWO HEME GROUPS AND COPPER B.
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- PTM: HIS-276 AND TYR-280 RESIDUES ARE INVOLVED IN THE FORMATION OF
 CC A COPPER-COORDINATED COVALENT CROSSLINK AT THE ACTIVE SITE OF THE
 CC CATALYTIC SUBUNIT 1.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y07533; CAA68821.1; -
 CC PIR: S08270; S08270.
 CC PDB: 1AR1; 11-FEB-98.
 CC InterPro: IPR000883; COX1.
 CC Pfam: PF00115; COX1; 1.
 CC PRINTS: PR01165; CYCOXIDASE1.
 CC PROSITE: PS00077; COX1; 1.
 CC Oxioreductase; Heme; Copper; Transmembrane; Inner membrane;
 CC Respiratory chain; Hydrogen ion transport; 3D-structure.
 CC
 CC DOMAIN 1 28
 CC TRANSMEM 29 59 I. PERIPLASMIC.
 CC TRANSMEM 60 82 II. PERIPLASMIC.
 CC TRANSMEM 83 120 III. CYTOPLASMIC.
 CC TRANSMEM 121 126 IV. CYTOPLASMIC.
 CC TRANSMEM 127 151 V. CYTOPLASMIC.
 CC TRANSMEM 152 176 VI. CYTOPLASMIC.
 CC TRANSMEM 177 206 VII. CYTOPLASMIC.
 CC TRANSMEM 207 217 VIII. CYTOPLASMIC.
 CC TRANSMEM 218 251 IX. CYTOPLASMIC.
 CC TRANSMEM 252 262 X. CYTOPLASMIC.
 CC TRANSMEM 263 299 XI. CYTOPLASMIC.
 CC TRANSMEM 300 303 XII. CYTOPLASMIC.
 CC TRANSMEM 304 319 XIII. CYTOPLASMIC.
 CC TRANSMEM 320 332 XIV. CYTOPLASMIC.
 CC TRANSMEM 333 364 XV. CYTOPLASMIC.
 CC TRANSMEM 365 369 XVI. CYTOPLASMIC.
 CC TRANSMEM 370 395 XVII. CYTOPLASMIC.
 CC TRANSMEM 396 404 XVIII. CYTOPLASMIC.
 CC TRANSMEM 405 437 XIX. CYTOPLASMIC.
 CC TRANSMEM 438 440 XX. CYTOPLASMIC.
 CC TRANSMEM 441 469 XXI. CYTOPLASMIC.
 CC TRANSMEM 470 478 XXII. CYTOPLASMIC.
 CC TRANSMEM 479 514 XXIII. CYTOPLASMIC.
 CC TRANSMEM 515 558 XXIV. CYTOPLASMIC.
 CC TRANSMEM 559 566 XXV. CYTOPLASMIC.
 CC TRANSMEM 567 576 XXVI. CYTOPLASMIC.
 CC TRANSMEM 577 580 XXVII. CYTOPLASMIC.
 CC TRANSMEM 581 584 XXVIII. CYTOPLASMIC.
 CC TRANSMEM 585 588 XXIX. CYTOPLASMIC.
 CC TRANSMEM 589 592 XXX. CYTOPLASMIC.
 CC TRANSMEM 593 596 XXXI. CYTOPLASMIC.
 CC TRANSMEM 597 599 XXXII. CYTOPLASMIC.
 CC TRANSMEM 600 603 XXXIII. CYTOPLASMIC.
 CC TRANSMEM 604 607 XXXIV. CYTOPLASMIC.
 CC TRANSMEM 608 611 XXXV. CYTOPLASMIC.
 CC TRANSMEM 612 615 XXXVI. CYTOPLASMIC.
 CC TRANSMEM 616 619 XXXVII. CYTOPLASMIC.
 CC TRANSMEM 620 623 XXXVIII. CYTOPLASMIC.
 CC TRANSMEM 624 627 XXXIX. CYTOPLASMIC.
 CC TRANSMEM 628 631 XL. CYTOPLASMIC.
 CC TRANSMEM 632 635 XLI. CYTOPLASMIC.
 CC TRANSMEM 636 639 XLII. CYTOPLASMIC.
 CC TRANSMEM 640 643 XLIII. CYTOPLASMIC.
 CC TRANSMEM 644 647 XLIV. CYTOPLASMIC.
 CC TRANSMEM 648 651 XLV. CYTOPLASMIC.
 CC TRANSMEM 652 655 XLVI. CYTOPLASMIC.
 CC TRANSMEM 656 659 XLVII. CYTOPLASMIC.
 CC TRANSMEM 660 663 XLVIII. CYTOPLASMIC.
 CC TRANSMEM 664 667 XLIX. CYTOPLASMIC.
 CC TRANSMEM 668 671 L. CYTOPLASMIC.
 CC TRANSMEM 672 675 LI. CYTOPLASMIC.
 CC TRANSMEM 676 679 LII. CYTOPLASMIC.
 CC TRANSMEM 680 683 LIII. CYTOPLASMIC.
 CC TRANSMEM 684 687 LIV. CYTOPLASMIC.
 CC TRANSMEM 688 691 LV. CYTOPLASMIC.
 CC TRANSMEM 692 695 LVI. CYTOPLASMIC.
 CC TRANSMEM 696 699 LVII. CYTOPLASMIC.
 CC TRANSMEM 700 703 LVIII. CYTOPLASMIC.
 CC TRANSMEM 704 707 LIX. CYTOPLASMIC.
 CC TRANSMEM 708 711 LX. CYTOPLASMIC.
 CC TRANSMEM 712 715 LXI. CYTOPLASMIC.
 CC TRANSMEM 716 719 LXII. CYTOPLASMIC.
 CC TRANSMEM 720 723 LXIII. CYTOPLASMIC.
 CC TRANSMEM 724 727 LXIV. CYTOPLASMIC.
 CC TRANSMEM 728 731 LXV. CYTOPLASMIC.
 CC TRANSMEM 732 735 LXVI. CYTOPLASMIC.
 CC TRANSMEM 736 739 LXVII. CYTOPLASMIC.
 CC TRANSMEM 740 743 LXVIII. CYTOPLASMIC.
 CC TRANSMEM 744 747 LXIX. CYTOPLASMIC.
 CC TRANSMEM 748 751 LXX. CYTOPLASMIC.
 CC TRANSMEM 752 755 LXXI. CYTOPLASMIC.
 CC TRANSMEM 756 759 LXXII. CYTOPLASMIC.
 CC TRANSMEM 760 763 LXXIII. CYTOPLASMIC.
 CC TRANSMEM 764 767 LXXIV. CYTOPLASMIC.
 CC TRANSMEM 768 771 LXXV. CYTOPLASMIC.
 CC TRANSMEM 772 775 LXXVI. CYTOPLASMIC.
 CC TRANSMEM 776 779 LXXVII. CYTOPLASMIC.
 CC TRANSMEM 780 783 LXXVIII. CYTOPLASMIC.
 CC TRANSMEM 784 787 LXXIX. CYTOPLASMIC.
 CC TRANSMEM 788 791 LXXX. CYTOPLASMIC.
 CC TRANSMEM 792 795 LXXXI. CYTOPLASMIC.
 CC TRANSMEM 796 799 LXXXII. CYTOPLASMIC.
 CC TRANSMEM 800 803 LXXXIII. CYTOPLASMIC.
 CC TRANSMEM 804 807 LXXXIV. CYTOPLASMIC.
 CC TRANSMEM 808 811 LXXXV. CYTOPLASMIC.
 CC TRANSMEM 812 815 LXXXVI. CYTOPLASMIC.
 CC TRANSMEM 816 819 LXXXVII. CYTOPLASMIC.
 CC TRANSMEM 820 823 LXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 824 827 LXXXIX. CYTOPLASMIC.
 CC TRANSMEM 828 831 LXXXX. CYTOPLASMIC.
 CC TRANSMEM 832 835 LXXXXI. CYTOPLASMIC.
 CC TRANSMEM 836 839 LXXXXII. CYTOPLASMIC.
 CC TRANSMEM 840 843 LXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 844 847 LXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 848 851 LXXXXV. CYTOPLASMIC.
 CC TRANSMEM 852 855 LXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 856 859 LXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 860 863 LXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 864 867 LXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 868 871 LXXXXX. CYTOPLASMIC.
 CC TRANSMEM 872 875 LXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 876 879 LXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 880 883 LXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 884 887 LXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 888 891 LXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 892 895 LXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 896 899 LXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 900 903 LXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 904 907 LXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 908 911 LXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 912 915 LXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 916 919 LXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 920 923 LXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 924 927 LXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 928 931 LXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 932 935 LXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 936 939 LXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 940 943 LXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 944 947 LXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 948 951 LXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 952 955 LXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 956 959 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 960 963 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 964 967 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 968 971 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 972 975 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 976 979 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 980 983 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 984 987 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 988 991 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 992 995 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 996 999 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1000 1003 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1004 1007 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1008 1011 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1012 1015 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1016 1019 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1020 1023 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1024 1027 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1028 1031 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1032 1035 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1036 1039 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1040 1043 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1044 1047 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1048 1051 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1052 1055 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1056 1059 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1060 1063 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1064 1067 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1068 1071 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1072 1075 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1076 1079 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1080 1083 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1084 1087 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1088 1091 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1092 1095 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1096 1099 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1100 1103 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1104 1107 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1108 1111 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1112 1115 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1116 1119 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1120 1123 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1124 1127 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1128 1131 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1132 1135 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1136 1139 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1140 1143 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1144 1147 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1148 1151 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1152 1155 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1156 1159 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1160 1163 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1164 1167 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1168 1171 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1172 1175 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1176 1179 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1180 1183 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1184 1187 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1188 1191 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1192 1195 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1196 1199 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1200 1203 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1204 1207 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1208 1211 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1212 1215 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1216 1219 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1220 1223 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1224 1227 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1228 1231 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1232 1235 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1236 1239 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1240 1243 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1244 1247 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1248 1251 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1252 1255 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1256 1259 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1260 1263 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1264 1267 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1268 1271 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1272 1275 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1276 1279 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1280 1283 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1284 1287 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1288 1291 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1292 1295 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1296 1299 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1300 1303 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1304 1307 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1308 1311 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1312 1315 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1316 1319 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1320 1323 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1324 1327 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1328 1331 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1332 1335 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1336 1339 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1340 1343 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1344 1347 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1348 1351 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1352 1355 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1356 1359 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1360 1363 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1364 1367 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1368 1371 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1372 1375 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1376 1379 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1380 1383 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1384 1387 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1388 1391 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1392 1395 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1396 1399 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1400 1403 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1404 1407 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1408 1411 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1412 1415 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1416 1419 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1420 1423 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1424 1427 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1428 1431 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1432 1435 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1436 1439 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1440 1443 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1444 1447 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1448 1451 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1452 1455 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1456 1459 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1460 1463 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1464 1467 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1468 1471 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1472 1475 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1476 1479 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1480 1483 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1484 1487 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1488 1491 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1492 1495 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1496 1499 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1500 1503 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1504 1507 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1508 1511 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1512 1515 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1516 1519 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1520 1523 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1524 1527 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1528 1531 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1532 1535 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1536 1539 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1540 1543 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1544 1547 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1548 1551 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1552 1555 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1556 1559 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1560 1563 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1564 1567 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1568 1571 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1572 1575 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1576 1579 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1580 1583 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1584 1587 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1588 1591 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1592 1595 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1596 1599 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1600 1603 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1604 1607 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1608 1611 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1612 1615 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1616 1619 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1620 1623 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1624 1627 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1628 1631 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1632 1635 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1636 1639 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1640 1643 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1644 1647 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1648 1651 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1652 1655 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1656 1659 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1660 1663 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1664 1667 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1668 1671 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1672 1675 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1676 1679 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1680 1683 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1684 1687 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1688 1691 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1692 1695 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1696 1699 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1700 1703 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1704 1707 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1708 1711 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1712 1715 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1716 1719 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1720 1723 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1724 1727 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1728 1731 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1732 1735 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1736 1739 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1740 1743 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1744 1747 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1748 1751 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1752 1755 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1756 1759 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1760 1763 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1764 1767 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1768 1771 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1772 1775 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1776 1779 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1780 1783 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1784 1787 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1788 1791 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1792 1795 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1796 1799 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1800 1803 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1804 1807 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1808 1811 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1812 1815 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1816 1819 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1820 1823 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1824 1827 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1828 1831 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1832 1835 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1836 1839 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1840 1843 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1844 1847 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1848 1851 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1852 1855 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1856 1859 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1860 1863 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1864 1867 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1868 1871 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1872 1875 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1876 1879 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1880 1883 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1884 1887 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1888 1891 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1892 1895 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1896 1899 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1900 1903 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1904 1907 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1908 1911 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1912 1915 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1916 1919 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1920 1923 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1924 1927 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1928 1931 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1932 1935 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1936 1939 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1940 1943 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1944 1947 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1948 1951 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1952 1955 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1956 1959 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 1960 1963 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 1964 1967 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 1968 1971 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 1972 1975 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 1976 1979 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 1980 1983 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 1984 1987 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 1988 1991 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 1992 1995 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 1996 1999 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 2000 2003 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 2004 2007 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 2008 2011 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 2012 2015 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 2016 2019 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 2020 2023 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 2024 2027 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 2028 2031 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 2032 2035 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 2036 2039 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 2040 2043 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 2044 2047 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 2048 2051 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 2052 2055 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 2056 2059 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 2060 2063 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 2064 2067 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 2068 2071 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 2072 2075 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 2076 2079 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 2080 2083 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 2084 2087 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 2088 2091 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 2092 2095 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 2096 2099 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 2100 2103 LXXXXXXXVIII. CYTOPLASMIC.
 CC TRANSMEM 2104 2107 LXXXXXXXIX. CYTOPLASMIC.
 CC TRANSMEM 2108 2111 LXXXXXXXX. CYTOPLASMIC.
 CC TRANSMEM 2112 2115 LXXXXXXXXI. CYTOPLASMIC.
 CC TRANSMEM 2116 2119 LXXXXXXXII. CYTOPLASMIC.
 CC TRANSMEM 2120 2123 LXXXXXXXIII. CYTOPLASMIC.
 CC TRANSMEM 2124 2127 LXXXXXXXIV. CYTOPLASMIC.
 CC TRANSMEM 2128 2131 LXXXXXXXV. CYTOPLASMIC.
 CC TRANSMEM 2132 2135 LXXXXXXXVI. CYTOPLASMIC.
 CC TRANSMEM 2136 2139 LXXXXXXXVII. CYTOPLASMIC.
 CC TRANSMEM 2140

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X62645; CAA44514.1; -
DR PIR: S20534; S20534.
DR HSSP: P98002; IAR1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1.1.
KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;
KW Hydrogen ion transport.
FT METAL 102 102 IRON (HEME A) (PROBABLE).
FT METAL 284 284 COPPER B (PROBABLE).
FT METAL 288 288 COPPER B (PROBABLE).
FT METAL 333 333 COPPER B (PROBABLE).
FT METAL 334 334 COPPER B (PROBABLE).
FT METAL 419 419 IRON (HEME A3) (PROBABLE).
FT METAL 421 421 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 565 AA; 62685 MW; A74738C138DA04E1 CRC64;

Query Match 81.8%; Score 2453; DB 1; Length 565;
Best Local Similarity 78.7%; Pred. No. 1.5e-135;
Matches 446; Conservative 50; Mismatches 59; Indels 12; Gaps 3;

QY 1 MADAAIHGHHKOGFTFMSTNKHDIQLVLAAGVFTSVLTFTVYRLEMDPG 60
DB 1 MADAAIHGHE-HDRRGFTFMSTNKHDIQLVLTGELGLISVAFVYRMEELWAPG 59
QY 61 VOYUCLGEAR-----LIADASQCTANGHLNMYVYRHLMEFYRGTALGCGF 110
DB 60 VOYUCLGEARGLVKGFGFOSLPSAVENCSTNGHLMVMTITGHILMFEVYRALTGCGF 119
QY 111 GNYLPLDIGADMAFPRMNNLSFMLETAGTAMGVASLFARFGGOLSGGVWLYRPLS 170
DB 120 GNYLPLDIGADMAFPRMNNLSFMLETAGTAMGVASLFARFGGOLSGGVWLYRPLS 179
QY 171 TREAGYSMDLAFVAVHSGASSIGALIMTTFINMRAPGTLHKVPLFSWIFITAMLI 230
DB 180 TSEGSYTDLAFVAVHSGASSIGALIMTTFINMRAPGTLHKVPLFSWIFITAMLI 239
QY 231 LLALPVLGATITMLTDNFGTFFNAGGDDPLXOHLMEFGHREYITILGFGIIS 290
DB 240 LLALPVLGATITMLTDNFGTFFNAGGDDPLXOHLMEFGHREYITILGFGIIS 299
QY 291 HVSFTSKRPVEGYLPRVYAVAGVGLFVYVWVHMYTVGMSLTOQSGFMLATVIAVPT 350
DB 300 HVSFTSKRPVEGYLPRVYAVAGVGLFVYVWVHMYTVGMSLTOQSGFMLATVIAVPT 359
QY 351 GIKIFSWATWGGSVKESPMIAFGFMELFTVGVGTGVLAAQGLDRAYHDTYVVAH 410
DB 360 GIKIFSWATWGGSVKESPMIAFGFMELFTVGVGTGVLAAQGLDRAYHDTYVVAH 419
QY 411 FHYVMSLGATIFAGIYFYVPRKFSGRAPRAKLLHFWTFEIGANTFFRQHLGQGM 470
DB 420 FHYVMSLGATIFAGIYFYVPRKFSGRAPRAKLLHFWTFEIGANTFFRQHLGQGM 479
QY 471 PRYIDYDEAFALNNKVSYSAGLAFASLEFFIYFYVTVAGRRETRPNWGEFADTLE 530
DB 480 PRYIDYDEAFALNNKVSYSAGLAFASLEFFIYFYVTVAGRRETRPNWGEFADTLE 538
QY 531 WTLPSPPAHFFELPKRSDMKHPSH 557
DB 539 WTLPSPPAHFFELPKRSDMKHPSH 565

RESULT 3

CX1A_PARDE
ID CX1A_PARDE STANDARD; PRT; 554 AA.
AC P08305;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I-ALPHA (EC 1.9.3.1) (CYTOCHROME AA3
SUBUNIT 1-ALPHA).
GN CYADI OR COI.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum rubrum group;
OC Paracoccus.
OX NCBI_TaxID=266;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SI657;
RA Raitio M., Jalli T., Saraste M.;
RT "Isolation and analysis of the genes for cytochrome c oxidase in
RT Paracoccus denitrificans."
RL EMBO J. 6:2825-2833(1987).
CC -1- FUNCTION: SUBUNIT I AND II FORM THE FUNCTIONAL CORE OF THE ENZYME
CC COMPLEX. ELECTRON OXIDATION IN CYTOCHROME C ARE TRANSFERRED VIA
CC HEME A AND CU(A) TO THE BINDING CENTER FORMED BY HEME A3 AND
CC CU(B). THIS CYTOCHROME C OXIDASE SHOWS PROTON PUMP ACTIVITY
CC ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS AND COPPER B.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X05829; CAA29274.1; -
DR PIR: S03809; S03809.
DR HSSP: P98002; IAR1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1.1.
KW Oxidoreductase; Heme; Copper; Transmembrane; Inner membrane;
KW Respiratory chain; Hydrogen ion transport.
FT METAL 91 91 IRON (HEME A) (BY SIMILARITY).
FT METAL 273 273 COPPER B (BY SIMILARITY).
FT METAL 277 277 COPPER B (BY SIMILARITY).
FT METAL 322 322 COPPER B (BY SIMILARITY).
FT METAL 323 323 COPPER B (BY SIMILARITY).
FT METAL 406 406 IRON (HEME A3) (BY SIMILARITY).
FT METAL 408 408 IRON (HEME A) (BY SIMILARITY).
FT DISULFID 63 77 BY SIMILARITY.
FT TRANSMEM 26 56 BY SIMILARITY.
FT TRANSMEM 81 118 BY SIMILARITY.
FT TRANSMEM 127 148 BY SIMILARITY.
FT TRANSMEM 175 203 BY SIMILARITY.
FT TRANSMEM 215 248 BY SIMILARITY.
FT TRANSMEM 260 295 BY SIMILARITY.
FT TRANSMEM 301 319 BY SIMILARITY.
FT TRANSMEM 331 359 BY SIMILARITY.
FT TRANSMEM 367 390 BY SIMILARITY.
FT TRANSMEM 399 425 BY SIMILARITY.
FT TRANSMEM 436 463 BY SIMILARITY.
FT TRANSMEM 478 508 BY SIMILARITY.
SQ SEQUENCE 554 AA; 62012 MW; 58AD591FBDCE794 CRC64;

Query Match

78.2%; Score 2347; DB 1; Length 554;

Best Local Similarity 76.1%; Pred. No. 2,1e-129;
Matches 415; Conservative 59; Mismatches 69; Indels 2; Gaps 1;

QY 13 EKGGFTTRPMSTNHRDGLLYVAAGVGTISVLTVMRLMDPQVQYMCBSARLI 72
DB 11 EKGGFTTRPMSTNHRDGLLYVAAGVGTISVLTVMRLMDPQVQYMCBSARLI 70
QY 73 ADASQTCNANGHLMNVMVYHGLIMFPGVIRALPGFCNLYLMPDLOGADPAFRPMNL 132
DB 71 ADAACCTPAHLMNVYVYHGLIMFPGVIRALPGFCNLYLMPDLOGADPAFRPMNL 130
QY 133 SFMLFAGTAMGVALSPAFGGDQGLSGVGVWVLPPLSTREAGYNDLAFVHLSGASS 192
DB 131 SYVLVCGVSLATSLSPGGSDQPGAGVGVWVLPPLSTREAGYNDLAFVHLSGASS 190
QY 193 IMCAIMWITTFELMRAPGKTLHKVPLFSWISITFAMLLALPVLGATITMLTDRENT 252
DB 191 ILCAIMWITTFELMRAPGKTLHKVPLFSWISITFAMLLALPVLGATITMLTDRENT 250
QY 253 TFFNPAGGGPILYOHILMFEGHPEVYIILPGFIIISHVSTFSKPKPVGYLPMYVAV 312
DB 251 QFFDPAGGGDPVLYOHILMFEGHPEVYIILPGFIIISHVSTFSKPKPVGYLPMYVAV 310
QY 313 AIGVLGFVVAHHMYTVGMSLTQOQSYFMLATVYIAVPTGIKIFSWIATMGGSVFEKSPM 372
DB 311 AIAFLGFVVAHHMYTVGMSLTQOQSYFMLATVYIAVPTGIKIFSWIATMGGSVFEKSPM 370
QY 373 LMAFGMPLFTVGVGVYLAQGLDRAYHDVYVVAHHMYTVGMSLTQOQSYFMLATVYIAVPTGIKIFSWIATMGGSVFEKSPM 432
DB 371 LMAFLA--FLETVGVGVYLAQGLDRAYHDVYVVAHHMYTVGMSLTQOQSYFMLATVYIAVPTGIKIFSWIATMGGSVFEKSPM 428
QY 433 KFSGRAPPEMAKILHWPFTIGANTVFPOHFLRGOMPRRYIDYDEAFALNKKVSSYGA 492
DB 429 KMSGRQYPRBAOGLHWMFMISGLNLFPPQHLRGOMPRRYIDYDEAFALNKKVSSYGA 488
QY 493 FLAFAFLFTFIVYTVLAGRRETRPNPGEFADTLENTLDSPPRAHFEETLPKRSMD 552
DB 489 YISFAFLFTFIVYTVLAGRRETRPNPGEFADTLENTLDSPPRAHFEETLPKRSMD 548
QY 553 KHRSH 557
DB 549 RAOAH 553

RESULT 4
COX1_RHILE STANDARD: PRT: 538 AA.
ID COX1_RHILE
AC 008855;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME AA3 SUBUNIT 1).
GN CTAD OR COXA.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=384.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94161496; PubMed=8117073;
RT Gabel C., Bittinger M.A., Maier R.J.;
RT "Cytochrome a3 gene regulation in members of the family
RT Rhizobiaceae: comparison of copper and oxygen effects in
RT Bradyrhizobium japonicum and Rhizobium tropici.";
RL Appl. Environ. Microbiol. 60:141-148(1994).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERROCYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 12
CC POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- DEVELOPMENTAL STAGE: FREE IN SOIL (NOT AS BACTERIUM).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X74341; CA52388.1; -.
CC HSSP: P98002; IARL.
CC InterPro: IPR000883; COX1.
CC Pfam: PF00115; COX1; 1.
CC PRINTS: PR01165; CYCOXIDASEI.
CC PROSITE: PS00077; COX1; 1.
CC Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
CC
CC TRANSMEM 37 57 POTENTIAL.
CC TRANSMEM 84 104 POTENTIAL.
CC TRANSMEM 121 141 POTENTIAL.
CC TRANSMEM 170 190 POTENTIAL.
CC TRANSMEM 208 228 POTENTIAL.
CC TRANSMEM 259 279 POTENTIAL.
CC TRANSMEM 291 311 POTENTIAL.
CC TRANSMEM 329 349 POTENTIAL.
CC TRANSMEM 362 382 POTENTIAL.
CC TRANSMEM 401 421 POTENTIAL.
CC TRANSMEM 438 458 POTENTIAL.
CC TRANSMEM 476 496 POTENTIAL.
CC METAL 82 82 IRON (HEME A) (PROBABLE).
CC METAL 265 265 COPPER B (PROBABLE).
CC METAL 269 269 COPPER B (PROBABLE).
CC METAL 314 314 COPPER B (PROBABLE).
CC METAL 315 315 COPPER B (PROBABLE).
CC METAL 400 400 IRON (HEME A3) (PROBABLE).
CC METAL 402 402 IRON (HEME A) (PROBABLE).
CC SEQUENCE 538 AA; 59133 MW; 740C8FEC902D9BF CRC64;
Query Match 62.5%; Score 1873.5; DB 1; Length 538;
Best Local Similarity 64.2%; Pred. No. 5.3e-102;
Matches 348; Conservative 63; Mismatches 110; Indels 21; Gaps 4;
QY 7 HGHDRHEKQGFTRPMSTNHRDGLLYVAAGVGTISVLTVMRLMDPQVQYMC 66
DB 15 HGHDRHAPR--TGMRRSTNHRDGLLYVAAGVGTISVLTVMRLMDPQVQYMC 69
QY 67 EGARLIADASQTCNANGHLMNVMVYHGLIMFPGVIRALPGFCNLYLMPDLOGADPAFR 126
DB 70 -----HNTILNVVFTVSHGVIMIFPMVPMAMGGGNNFLPLMIGAPMAF 115
QY 127 PRMNLSPMLFAGTAMGVALSPAFGGDQGLSGVGVWVLPPLSTREAGYNDLAFVHLS 185
DB 116 PRMNLSPMLFAGTAMGVALSPAFGGDQGLSGVGVWVLPPLSTREAGYNDLAFVHLS 175
QY 186 HSGASSITGAINMTTFELMRAPGKTLHKVPLFSWISITFAMLLALPVLGATITML 245
DB 176 HSGASSITGAINMTTFELMRAPGKTLHKVPLFSWISITFAMLLALPVLGATITML 235
QY 246 TDRNFTTFFNPAGGGDPVLYOHILMFEGHPEVYIILPGFIIISHVSTFSKPKPVGYL 305
DB 236 TDRNFTTFFNPAGGGDPVLYOHILMFEGHPEVYIILPGFIIISHVSTFSKPKPVGYL 295
QY 306 PMYVMAVAGVGLFVVAHHMYTVGMSLTQOQSYFMLATVYIAVPTGIKIFSWIATMGG 365
DB 296 GMYVMAVAGVGLFVVAHHMYTVGMSLTQOQSYFMLATVYIAVPTGIKIFSWIATMGG 355

QY 366 VEFKSPMLMAFGFMEFLFTVGGTGVLAQGLDRAYHDTYVVAHHFYVMSLGAIFAIFA 425
 DB 356 IEFAPMLMALAFLEFLFTVGGTGVLANASLDRVLHDTYVVAHHFYVLSLGAIFAIFA 415
 QY 426 GIYFMPKESGAFPEMAAKLHFWTEFFIGANTFPPHFLGQGMRRYIDYPEAFALMN 485
 DB 416 GWYFMPKESGAFPEMAAKLHFWTEFFIGANTFPPHFLGQGMRRYIDYPEAFALMN 475
 QY 486 KVSXGAFPLAFASFLFTVYTLVAGRRTRPMPGCEPADTLEWTLPSPPAHFEPL 545
 DB 476 LVSSIGSVISGVSFLFTYC-VYDAFAKNVPYGDNPWGAGATTLLEWTLPSPPAHFEPL 534
 QY 546 PK 547
 DB 535 PR 536

RESULT 5
 COX1_BRAJA STANDARD; PRT: 541 AA.
 AC P31833;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME AA3
 SUBUNIT 1).
 GN CTAD OR COXA.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LO;
 RX MEDLINE=91045095; PubMed=2172930;
 RA Gabel C., Maier R.J.;
 RT "Nucleotide sequence of the coxa gene encoding subunit I of
 RT cytochrome a3 of Bradyrhizobium japonicum.";
 RL Nucleic Acids Res. 18:6143-6143(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-USA 110SPC4;
 RX MEDLINE=91211625; PubMed=1965217;
 RA Bock M., Bolliger M., Hennecke H.;
 RT "Genetic analysis of the cytochrome c-a3 branch of the
 RT Bradyrhizobium japonicum respiratory chain.";
 RL Mol. Microbiol. 4:2147-2157(1990).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERROCYTOCHROME C.
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 12
 CC POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- DEVELOPMENTAL STAGE: FREE IN SOIL (NOT AS BACTERIUM).
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC DR EMBL; X54318; CAA38216.1; -;
 CC DR EMBL; X54800; CAA38570.1; -;
 CC PIR; S13076; ODZJ1.

DR HSSP; P00396; 10CC.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1; 1.
 DR PRINTS; PRO1165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
 FT TRANSMEM 40 60
 FT TRANSMEM 87 107
 FT TRANSMEM 124 144
 FT TRANSMEM 173 193
 FT TRANSMEM 211 231
 FT TRANSMEM 262 282
 FT TRANSMEM 294 314
 FT TRANSMEM 332 352
 FT TRANSMEM 365 385
 FT TRANSMEM 404 424
 FT TRANSMEM 441 461
 FT TRANSMEM 483 503
 FT METAL 85 85
 FT METAL 268 268
 FT METAL 272 272
 FT METAL 317 317
 FT METAL 318 318
 FT METAL 403 403
 FT METAL 405 405
 SO SEQUENCE 541 AA; 59265 MW; 08F9A69D0C90DB85 CRC64;

Query Match 61.0%; Score 1829; DB 1; Length 541;
 Best Local Similarity 61.8%; Pred. No. 2e-99;
 Matches 337; Conservative 72; Mismatches 112; Indels 24; Gaps 4;

QY 7 HGHHEKQGFHFMSTNHNKIDGLYLVAAGVISTFLFYVMELEMDQGYMCL 66
 DB 15 HGHDEHHPGWRVYSTNHNKIDGLYLVAAGVISTFLFYVMELEMDQGYMCL 72
 QY 67 EGARLADASOTYCTANGHLMNVVYTHGILMFEFVGIPALFGSGNYLMLQIGAPDMAF 126
 DB 73 -----HETHYVNFVTSGLIMFEFVGIPALFGSGNYLMLQIGAPDMAF 118
 QY 127 PRMNLSPWFLFIAGTAGVSLPAPGDSGVGVLYLPPLSTR-EGYSDMLAFAY 185
 DB 119 PRMNLSPWFLFIAGTAGVSLPAPGDSGVGVLYLPPLSTR-EGYSDMLAFAY 178
 QY 186 HLSGASSIMGAINITFELNMRAPGMLHVRPLFSNISFTLMLLILALPYLAGATMLL 245
 DB 179 HLAGASSILGAINITFELNMRAPGMLHVRPLFSNISFTLMLLILALPYLAGATMLL 238
 QY 246 TDRNFGTFEPNAGGDPILYOHILMFEFGHEVYIILPFGIISHVYSTSKRPVGYL 305
 DB 239 TDRNFGTFEPNAGGDPILYOHILMFEFGHEVYIILPFGIISHVYSTSKRPVGYL 298
 QY 306 PMYAMVAIVGLGVVAHNMVYVAGSLTQOSTFMATVYIAPVPTGKIFSWATMGGS 365
 DB 299 GMAYAMVAIVGLGVVAHNMVYVAGSLTQOSTFMATVYIAPVPTGKIFSWATMGGS 358
 QY 366 VEFKSPMLMAFGFMEFLFTVGGTGVLAQGLDRAYHDTYVVAHHFYVMSLGAIFAIFA 425
 DB 359 IEFAPMLMALAFLEFLFTVGGTGVLANASLDRVLHDTYVVAHHFYVLSLGAIFAIFA 418
 QY 426 GIYFMPKESGAFPEMAAKLHFWTEFFIGANTFPPHFLGQGMRRYIDYPEAFALMN 485
 DB 419 GWYFMPKESGAFPEMAAKLHFWTEFFIGANTFPPHFLGQGMRRYIDYPEAFALMN 478
 QY 486 KVSXGAFPLAFASFLFTVYTLVAGRRTRPMPGCEPADTLEWTLPSPPAHFEPL 542
 DB 479 LVSSIGSVISG-----FVLLFTYLCVDAFAKNVPYGDNPWGAGATTLLEWTLPSPPAHFEPL 534
 QY 543 ETLPR 547
 DB 535 EVLPR 539

RESULT 6
COX1_MARPO STANDARD: PRT: 522 AA.
AC P26856;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Marchantia polymorpha (liverwort).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiales; Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114051; PubMed=1731062;
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;
RT "Gene organization deduced from the complete sequence of liverwort
Marchantia polymorpha mitochondrial DNA. A primitive form of plant
mitochondrial genome.";
RT J. Mol. Biol. 223:1-7(1992).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL
INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
or send an email to license@sdb-sb.ch).
CC EMBL: M68929; AAC09451.1; -
DR PIR: S25956; S25956.
DR HSP: P00396; 10CC.
DR Wende; 2055; MARPO; cox1.1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 65 65 IRON (HEME A) (PROBABLE).
FT METAL 244 244 COPPER B (PROBABLE).
FT METAL 248 248 COPPER B (PROBABLE).
FT METAL 293 293 COPPER B (PROBABLE).
FT METAL 294 294 COPPER B (PROBABLE).
FT METAL 379 379 IRON (HEME A3) (PROBABLE).
FT METAL 381 381 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 522 AA; 57551 MW; 4B88E16F03A11D01 CRC64;

Query Match 58.5%; Score 1756.5; DB 1; Length 522;
Best Local Similarity 58.9%; Pred. No. 3.1e-95;
Matches 315; Conservative 92; Mismatches 101; Indels 27; Gaps 6;

OY 17 FTRFWMSTNHRKDIGLTYAACVGFISLFTVYMRLEMDGQVGYMCEGRLADAS 76
DB 4 FAORWLFSTNHRKDIGLTYAACVGFISLFTVYMRLEMDGQVGYMCEGRLADAS 76
N 48

OY 77 QTCANGHLNWNVYTHGILMFEVFGIPALFGSGNLYLMDIGADPAPRRNNLSFWL 136
DB 49 QILGNGHQLVVLITAAHAFLEFFWVPMAMIGCGNMFVILIGSPMAPRRLNINISFWL 108
OY 137 FIAGTACVASFAPGGDQIGSGVGLVPLPSL--TREACYSMDLAFVAVHLSGASSIM 134
DB 109 LPPSLILLSSALV----EVGCGSGWTVYPPPLSGITSHSGSVDLAIFSLHSGVSSIL 163
OY 195 GAINMTTFELNMRAPGMDLHKVPLPFSMIFITAVLILALPVLAGATMTLDRNFGTTF 254
DB 164 GSINFTITTFMRAPGLTMHRLPLFVMSVLTATLILSLPVLGATLMTLDRNFTTF 223
OY 255 FNPAGGGDPLLYOHLMEFGHPEVYIILPFGIISHVSTFSKRPVGYLPMYAVAAV 314
DB 224 FDPAGGGDPLLYOHLMEFGHPEVYIILPFGIISHVSTFSKRPVGYLPMYAVAAV 283
OY 315 GVLGFVVAHHMTYVNGSLTQGSFMTATMYIANPTGIRKIFSMATMWSGVSKSPML 374
DB 284 GVLGIVVAHHMTYVNGSLTQGSFMTATMYIANPTGIRKIFSMATMWSGVSKSPML 343
OY 375 AEGFMEFTVGVTGIVLAQAGLDRAYHDTYVVAHHFVWMSLGAIFAIFAGIFYMPKF 434
DB 344 AVGFTEFTVGGLTIGIVLANSQVDIALHDITYVVAHHFVWMSLGAIFAIFAGIFYMPKF 403
OY 435 SGRAPEMAAKLHFWTEFIGNVTFEPQHLFGROGMRRYIDYPEAFLNNKVSYSYGL 494
DB 404 TGLQPEFTLQIHFWTFEFGVNLTFEPHFLGLAGMRPRIDYDAVAGNNAFSSFSYV 463
OY 495 AFAS-LEFFIYIFVYTLVAGRRTRPNWG--EFADLEWTLSPPPAHFFETLP 546
DB 464 SVGIFCEFFVVFLL--TLTSNKCASPFAVEONSTLEWVPSPPAHFFETLP 516

RESULT 7
COX1_BETVU STANDARD: PRT: 524 AA.
AC P24794;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Beta vulgaris (Sugar beet).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Caryophyllales; Magnoliophyta; eudicotyledons; core eudicots;
OX NCBI_TaxID=3555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TK81-0;
RX MEDLINE=9130331; PubMed=1651175;
RA Senda M., Harada T., Mikami T., Sugita M., Kinoshita T.;
RT "Genomic organization and sequence analysis of the cytochrome oxidase
subunit II gene from normal and male-sterile mitochondria in sugar
beet.";
RT Curr. Genet. 19:175-181(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Harada T., Mikami T., Kinoshita T.;
RT "Nucleotide sequence of cytochrome c oxidase subunit I gene of sugar
beet mitochondria.";
RT Proc. Sugar Beet Res. 29:15-21(1987).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57693; CAA40874.1; -;
CC EMBL: M57645; AAA87330.1; -;
CC HSSP: P00396; LOCC.
CC Mendel: 5481; BETVU:cox1.1.
CC InterPro: IPR000883; COX1.
CC Pfam: PF00115; COX1; 1.
CC PRINTS: PR01165; CYCOXIDASE1.
CC PROSITE: PS00077; COX1; 1.
CC Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
CC Respiratory chain; Inner membrane.
CC METAL 64 64 IRON (HEME A) (PROBABLE).
CC METAL 243 243 COPPER B (PROBABLE).
CC METAL 247 247 COPPER B (PROBABLE).
CC METAL 292 292 COPPER B (PROBABLE).
CC METAL 293 293 COPPER B (PROBABLE).
CC METAL 378 378 IRON (HEME A3) (PROBABLE).
CC METAL 380 380 IRON (HEME A) (PROBABLE).
CC CONFLICT 2 9 INLWLE -> AV (IN REF. 2).
CC CONFLICT 220 226 TTFEPA -> REFLFW (IN REF. 2).
CC CONFLICT 229 229 MISSING (IN REF. 2).
CC CONFLICT 501 503 EMM -> ND (IN REF. 2).
CC SEQUENCE 524 AA; 57580 MW; A862089E5C476EA CRC64;
Query Match 58.38; Score 1748.5; DB 1; Length 524;
Best Local Similarly 59.28; Pred. No. 9e-95;
Matches 315; Conservative 86; Mismatches 106; Indels 25; Gaps 5;
OY 20 RMPSTNKHDKGLLYVAGVGVFLTYMRLMDPVOYMCLEGARLIDAQOTC 79
DB 6 RMLSTNKHDKGLTYLFTFGATAGWGCFSVLMELARPE-----DQIL 50
OY 80 TANGHNVNVTYHGLIMPFVGIAPALFGSGFNGYLMPLQIGAPMAPRNNLSFMLEIA 139
DB 51 GGNHQLVNLTAHAPLMEFVMWPMAMIGSGNMFVILICAPMAPRNLNINISFWLLP 110
OY 140 GTANGVASLFRPGDGLGSGVNVLPPLS--TREAGYSMDLAFVHUSGASSINGAI 197
DB 111 SLILLSSALV-----EVGSGTGWTVYRPLSGITSHSGGAVDLAIFSIHLSGVSSILGSI 165
OY 198 NMITFLMRAKPMGLKRLKVPLEFSIFITAWLILALPYLAGATLMTLTDNFGTFENP 257
DB 166 NFITTFEMRGRPMTHRLPLTFVMSVLTATLALLSLPVLGATLMTLTDNENFTTFDP 225
OY 258 AGGDPLLYQHLLMEFGHREYVYLLPFGIISHVSTFSKRPFGYLPWYVAVAIGVL 317
DB 226 AGCGPILYQHLMFEGHREYVYLLPFGIISHVSTFSKRPFGYLPWYVAVAIGVL 285
OY 318 GFVVAHMHMTYVGSLSLQOSTFMATLVIANVPTGKIFSKATMWSGVFEKSPMLNAG 377
DB 286 GFVVAHMHMTYVGSLSLQOSTFMATLVIANVPTGKIFSKATMWSGVFEKSPMLNAG 345
OY 378 FMFLTYGVGTGIVLAAGLDRAYHDYVVAHFNHYMSLGAFAFAGTIFYVYKPSGR 437
DB 346 FFLFTYVGLTGIYVLAAGLDRAYHDYVVAHFNHYMSLGAFAFAGTIFYVYKPSGR 405
OY 438 AFPEMAALHFWTEFGANVTFFPOHFLGROGMPRRYIDYDEAFALNNKVSYGAFLAFA 497
DB 406 TYPELIGQIHFWITFFGVNLTFEPMHFLGLSGMPRRIPDYDAVAGNNALSSFGSYISV 465
OY 498 SFLPFIVLVYTVAGR-RETRPRNPG--ERADTLLEMTLPSPPAHFFETLP 546

DB 466 GICCFVWVITLTSKKNRCAPSPWAVEENSTYLLMMVWSPPAFTFGLP 517
RESULT 8
AC COX1_RICPR STANDARD; PRT; 534 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME
DE A3 SUBUNIT 1).
GN CTAD OR COXA OR RP405.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RA "The bacterial origin of mitochondria inferred from a phylogenetic
RT analysis of the cytochrome b and cytochrome c oxidase I genes.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RA MEDLINE=9035949; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Scharitz-Ponten T., Almark U.C.M., Podowski R.M., Neeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria".
RL Nature 396:133-140(1998).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y13859; CAA74167.1; -;
CC EMBL: AJ235271; CAA14862.1; -;
CC InterPro: IPR000883; COX1.
CC Pfam: PF00115; COX1; 1.
CC PRINTS: PR01165; CYCOXIDASE1.
CC PROSITE: PS00077; COX1; 1.
CC Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;
CC Complete proteome.
KW TRANSMEM 35 55 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 140 140 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 320 340 POTENTIAL.


```

OY 138 NMITTEFLNMRAPGMLTKHVPLEFSWISITAMILLALPVLGATIMLLDRNFETTFEPN 257
D 166 NFETITFNMKRGPMTHRLPLFVMSVLVTAFLLLSLPVLGATIMLLDRNFETTFEPN 225
OY 258 AGGSDPILYOHILMFEGHPEVYIIILPGFIIISHVSTFSKRPVEGYLPMYAMAIVGL 317
D 226 AGGSDPILYOHILMFEGHPEVYIIILPGFIIISHVSTFSKRPVEGYLPMYAMAIVGL 285
OY 318 GFVVAHMHMTYVMSLTOOSYEMLATVIAVPTGKIFSWATIMMGSEVERKSPMLAMG 377
D 286 GFVVAHMHMTYVMSLTOOSYEMLATVIAVPTGKIFSWATIMMGSEVERKSPMLAMG 345
OY 378 FMFLFTVGVGTGIVLAQAQGLDRAVHDYVVVAHFHVMSLGAIFAFAIGFYMPKESGR 437
D 346 FIFLFTIGLIGIVLANSGDIALHDYVVVAHFHVMSLGAIFAFAIGFYMPKESGR 405
OY 438 AFPEMAAKLHMTFFPGANVTFPPQHFGRGMPRRYIDYEAFALMKNVSYGAFLAFA 497
D 406 TYPELIGIHFMTFFGVNLTFFPMHFLGSLGMPRRIPDYDAAGMVALSSFGSYISV 465
OY 498 SFLEFIVFYVTLVAG--RETREPNG--EF-ADTLEMTLPSPPAHTEPLP 546
D 466 GICCFVVAITLSSGNKRCAPSPMAVEONPTTLEMLVOSPAPAFHTFGLP 517

RESULT 10
COX1_ORYSA
ID COX1_ORYSA STANDARD; PRT: 524 AA.
AC P14578;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Oryza sativa (Rice).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ethnaloideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90016815; PubMed=2552410;
RA Kadowaki K., Suzuki T., Kazama S., Oh-Fuchi T., Sakamoto W.;
RT "Nucleotide sequence of the cytochrome oxidase subunit I gene from
RT rice mitochondria."
RL Nucleic Acids Res. 17:7519-7519(1989).
RN [2]
RP SEQUENCE OF 1-57 FROM N.A.
RC STRAIN=CV, TAICHUNG 65; TISSUE=Shoot;
RX MEDLINE=92035104; PubMed=1718614;
RA Suzuki T., Kazama S., Hirai A., Akihama T., Kadowaki K.;
RT "The rice mitochondrial nad3 gene has an extended reading frame at
RT its 5' end: nucleotide sequence analysis of rice trns, nad3, and
RT rps12 genes."
RL Curr. Genet. 20:331-337(1991).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC 4 FERRICCYTOCHROME C.
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15990; CA34122.1; -
DR EMBL: M57903; AA70312.1; -
DR PIR: S06761; ODR21.
DR HSSP: P00396; 10CC.
DR Mendel: 5485; Oryza:cox1.1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.
DR PRINTS: PR01165; CYCOXIDASEI.
DR PROSITE: PS00077; COX1. 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 64 64 IRON (HEME A) (PROBABLE).
FT METAL 243 243 COPPER B (PROBABLE).
FT METAL 247 247 COPPER B (PROBABLE).
FT METAL 292 292 COPPER B (PROBABLE).
FT METAL 293 293 COPPER B (PROBABLE).
FT METAL 378 378 IRON (HEME A3) (PROBABLE).
FT METAL 380 380 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 524 AA; 57767 MW; F769B9E7106724E0 CRC64;

Query Match 57.7%; Score 1730.5; DB 1; Length 524;
Best Local Similarity 58.5%; Pred. No. 9.9e-94;
Matches 311; Conservative 88; Mismatches 108; Indels 25; Gaps 5;

OY 20 RWFMTNHRKDIGILYVAAGVGFISLFTYRMLTDPGQYMCLEGARLIADASQTC 79
D 6 RWFSTNHRKDIGILYFISGAIAGVMTGCFVLLRMELARG-----DQLL 50
OY 80 TANGHLMNVVYTHGILMNFVFGIPALPFGFGNYLMPLOIGADMAFPMMNLSEWLFIA 139
D 51 GGNHOLYVNLITAHAFMLTFVWPMAMTGFQWFWPIIIGADMAFPRLNNISFWLLP 110
OY 140 GTFMAGVASLEFAPGDGLSGVGVNLYPLPS--TRAGYSMDLAFVHLSGASSTMGAI 197
D 111 SLILLSSALV-----EVSGTGTWYVPLSGITSHSGGAVDLALPSLHSGVSSILGSI 165
OY 198 NMITTEFLNMRAPGMLTKHVPLEFSWISITAMILLALPVLGATIMLLDRNFETTFEPN 257
D 166 NFETITFNMKRGPMTHRLPLFVMSVLVTAFLLLSLPVLGATIMLLDRNFETTFEPN 225
OY 258 AGGSDPILYOHILMFEGHPEVYIIILPGFIIISHVSTFSKRPVEGYLPMYAMAIVGL 317
D 226 AGGSDPILYOHILMFEGHPEVYIIILPGFIIISHVSTFSKRPVEGYLPMYAMAIVGL 285
OY 318 GFVVAHMHMTYVMSLTOOSYEMLATVIAVPTGKIFSWATIMMGSEVERKSPMLAMG 377
D 286 GFVVAHMHMTYVMSLTOOSYEMLATVIAVPTGKIFSWATIMMGSEVERKSPMLAMG 345
OY 378 FMFLFTVGVGTGIVLAQAQGLDRAVHDYVVVAHFHVMSLGAIFAFAIGFYMPKESGR 437
D 346 FIFLFTIGLIGIVLANSGDIALHDYVVVAHFHVMSLGAIFAFAIGFYMPKESGR 405
OY 438 AFPEMAAKLHMTFFPGANVTFPPQHFGRGMPRRYIDYEAFALMKNVSYGAFLAFA 497
D 406 TYPELIGIHFMTFFGVNLTFFPMHFLGSLGMPRRIPDYDAAGMVALSSFGSYISV 465
OY 498 SFLEFIVFYVTLVAG--RETREPNG--EF-ADTLEMTLPSPPAHTEPLP 546
D 466 GICCFVVAITLSSGNKRCAPSPMAVEONPTTLEMLVOSPAPAFHTFGLP 517

RESULT 11
ID COX1_MAIZE STANDARD; PRT: 528 AA.
AC P08742;
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```



```

QY 20 RMEFSTNHKDIGLLYLVAAGVGFISVLTFTVYKRELMDPGVQVWCLLEGARLIADASQTC 79
    ||||||| : : : : : |||
    6 RMEFSTNHKDIGLLYLVAAGVGFISVLTFTVYKRELMDPGVQVWCLLEGARLIADASQTC 50
    ||||||| : : : : : |||
QY 80 TANGHLMNNVWVYTHGILMFMFGVCPALFGFGVYIMPLDQIGAPDMAFPRNNISFWLFTA 139
    ||||||| : : : : : |||
    51 GGNHQLYNLITAHALMFMFMFGVCPALFGFGVYIMPLDQIGAPDMAFPRNNISFWLFTA 110
    ||||||| : : : : : |||
QY 140 GTAMVASLFAFGDGLSAGVGVLYPLST--TREAGYSMDLAIFAVALSGASSIMGAI 197
    : : : : : ||| : : : : : |||
    111 SLLLSLSALV-----EVSGTGMVYPPPLSGTSSGCAVDLAIFSLSLSISLTSI 165
    ||||||| : : : : : |||
QY 198 NMITTELMNMRAGMTLHKVPLFSWSTFTAWMLLLAPVLGATIMLLTDNRNFTFENE 257
    ||||||| : : : : : |||
    166 NMTTTFNMRGPMTHMRPLFLFWSVLAFLFLLSLPLVALGATIMLLTDNRNFTFENE 225
    ||||||| : : : : : |||
QY 258 AGGDDLLVONHLMFEGHREVVYIILPGGIIISHVSTFSKRPVSGYLPVYVAMAIGVL 317
    ||||||| : : : : : |||
    226 AGGDDLLVONHLMFEGHREVVYIILPGGIIISHVSTFSKRPVSGYLPVYVAMAIGVL 285
    ||||||| : : : : : |||
QY 318 GFVYVMAHMYTVGMSLTQOSYFMLATVAVPTGKIFSMIATMNGSVEFKSPMLMAG 377
    ||||||| : : : : : |||
    286 GFVYVMAHMYTVGMSLTQOSYFMLATVAVPTGKIFSMIATMNGSVEFKSPMLMAG 345
    ||||||| : : : : : |||
QY 378 FMEFTVGVGTGIVLAOAGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRKFSGR 437
    ||||||| : : : : : |||
    346 FMEFTVGVGTGIVLAOAGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRKFSGR 405
    ||||||| : : : : : |||
QY 438 APPEMAKLHFMFTFEGANVTFFOHFLRGOGMPRXYIDYPEAFALMNVSSYGATLFA 497
    ||||||| : : : : : |||
    406 TPEFTLQOHFMFTFEGANVTFFOHFLRGOGMPRXYIDYPEAFALMNVSSYGATLFA 465
    ||||||| : : : : : |||
QY 498 SFLEFVIFVYTVLAGR-RETREPWG--EFADTLEMTLSPPAHTEFLP 546
    ||||||| : : : : : |||
    466 GIRFVVAITSSGKNCKACSPMAVBNPTLEMTLSPPAHTEFLP 517
    ||||||| : : : : : |||
Db 466 GIRFVVAITSSGKNCKACSPMAVBNPTLEMTLSPPAHTEFLP 517
    ||||||| : : : : : |||
RESULT 14
COX1_PROWI STANDARD: PRT: 514 AA.
AC 005143;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1 OR COXI.
OS Prototheca wickerhamii.
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
    Chlorellaceae; Prototheca.
OX NCBI_TaxID=3111;
RN 11]
RP SEQUENCE FROM N.A.
RX MEDLINE=93181269; PubMed=7680126;
    Wolff G., Burger G., Lang F.B., Kneek U.;
    "Mitochondrial genes in the colourless alga Prototheca wickerhamii
    resemble plant genes in their exons but fungal genes in their
    introns.";
    RT Nucleic Acids Res. 21:719-726(1993).
    RL Nucleic Acids Res. 21:719-726(1993).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
    CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
    3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
    CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
    CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
    AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
    AND COPPER B.
    CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
    4 FERROCYTOCHROME C.
    CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
    CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
    INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
    CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
    CC 1

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68721; CAA8661.1; -.
DR PIR: S30291; S30291.
DR HSSD: P00396; 10CC.
DR Mendel: 12219; PROWI; COX1.1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
    Respiratory chain; Inner membrane.
KW METAL 62 62 IRON (HEME A) (PROBABLE).
FT METAL 240 240 COPPER B (PROBABLE).
FT METAL 244 244 COPPER B (PROBABLE).
FT METAL 289 289 COPPER B (PROBABLE).
FT METAL 290 290 COPPER B (PROBABLE).
FT METAL 375 375 IRON (HEME A3) (PROBABLE).
FT METAL 377 377 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 514 AA; 56874 MW; C5219A94E8FB22 CRC64;

Query Match 57.1%; Score 1713.5; DB 1; Length 514;
Best local similarity 57.8%; Pred. No. 9.4e-93;
Matches 308; Conservative 91; Mismatches 105; Indels 29; Gaps 7;

QY 19 TRMEFSTNHKDIGLLYLVAAGVGFISVLTFTVYKRELMDPGVQVWCLLEGARLIADASQTC 78
    ||||||| : : : : : |||
    3 TRMEFSTNHKDIGLLYLVAAGVGFISVLTFTVYKRELMDPGVQVWCLLEGARLIADASQTC 47
    ||||||| : : : : : |||
QY 79 CTANGHLMNNVWVYTHGILMFMFGVCPALFGFGVYIMPLDQIGAPDMAFPRNNISFWLFTA 138
    ||||||| : : : : : |||
    48 LNSHQLYNLITAHALMFMFMFGVCPALFGFGVYIMPLDQIGAPDMAFPRNNISFWLFTA 107
    ||||||| : : : : : |||
QY 139 AGTAMVASLFAFGDGLSAGVGVLYPLST--REAGYSMDLAIFAVALSGASSIMGAI 196
    ||||||| : : : : : |||
    108 PSLLLSALV-----EVGSGTGMVYPPPLSGTSSGCAVDLAIFSLSLSISLTSI 162
    ||||||| : : : : : |||
QY 197 NMITTELMNMRAGMTLHKVPLFSWSTFTAWMLLLAPVLGATIMLLTDNRNFTFENE 236
    ||||||| : : : : : |||
    163 NMITTELMNMRAGMTLHKVPLFSWSTFTAWMLLLAPVLGATIMLLTDNRNFTFENE 221
    ||||||| : : : : : |||
QY 257 PAGGDDLLVONHLMFEGHREVVYIILPGGIIISHVSTFSKRPVSGYLPVYVAMAIGVL 316
    ||||||| : : : : : |||
    222 PAGGDDLLVONHLMFEGHREVVYIILPGGIIISHVSTFSKRPVSGYLPVYVAMAIGVL 281
    ||||||| : : : : : |||
QY 317 GFVYVMAHMYTVGMSLTQOSYFMLATVAVPTGKIFSMIATMNGSVEFKSPMLMAG 376
    ||||||| : : : : : |||
    282 GFVYVMAHMYTVGMSLTQOSYFMLATVAVPTGKIFSMIATMNGSVEFKSPMLMAG 341
    ||||||| : : : : : |||
QY 377 FMEFTVGVGTGIVLAOAGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRKFSGR 436
    ||||||| : : : : : |||
    342 FMEFTVGVGTGIVLAOAGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRKFSGR 401
    ||||||| : : : : : |||
QY 437 APPEMAKLHFMFTFEGANVTFFOHFLRGOGMPRXYIDYPEAFALMNVSSYGATLFA 496
    ||||||| : : : : : |||
    402 APPEMAKLHFMFTFEGANVTFFOHFLRGOGMPRXYIDYPEAFALMNVSSYGATLFA 461
    ||||||| : : : : : |||
QY 497 SFLEFVIFVYTVLAGR-RETREPWG--EFADTLEMTLSPPAHTEFLP 545
    ||||||| : : : : : |||
    462 SFLEFVIFVYTVLAGR-RETREPWG--EFADTLEMTLSPPAHTEFLP 512
    ||||||| : : : : : |||
Db 462 SFLEFVIFVYTVLAGR-RETREPWG--EFADTLEMTLSPPAHTEFLP 512
    ||||||| : : : : : |||
RESULT 15
COX1_CHOOR STANDARD: PRT: 532 AA.
AC P48866;

```


DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
 GN COX1.
 OS Chondrus crispus (Carragheen).
 OG Mitochondrion.
 OC Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartineaceae;
 OC Chondrus.
 CC NCBI_JtaxID=2769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Apices;
 RX MEDLINE=95341681; PubMed=7616569;
 RA Leblanc C., Boyen C., Richard O., Bonnard G., Grienenberger J.M.,
 RA Klarg B.;
 RT "Complete sequence of the mitochondrial DNA of the rhodophyte
 RT Chondrus crispus (Gigartinales). Gene content and genome
 RT organization.";
 RL J. Mol. Biol. 250:484-495(1995).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
 CC 4 FERROCYTOCHROME C.
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; 247547; CAAB7603.1; -;
 DR HSSP; P00396; 10CC.
 DR Mendel; 7724; CHOCr:cox1.1.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 69 69 IRON (HEME A) (PROBABLE).
 FT METAL 248 248 COPPER B (PROBABLE).
 FT METAL 252 252 COPPER B (PROBABLE).
 FT METAL 297 297 COPPER B (PROBABLE).
 FT METAL 298 298 COPPER B (PROBABLE).
 FT METAL 383 383 IRON (HEME A3) (PROBABLE).
 FT METAL 385 385 IRON (HEME A3) (PROBABLE).
 SEQUENCE 532 AA; 59290 MW; ADF18B4B35BA636F CRC64;

Query Match 55.9%; Score 1677.5; DB 1; Length 532;
 Best Local Similarity 56.1%; Pred. No. 1.2e-90;
 Matches 306; Conservative 94; Mismatches 110; Indels 35; Gaps 7;

QY 15 QGEFF-----RWMSTNNHNDIGLLTVAGVGFISLVTVYRLMLDPGVYKCLEGAR 70
 DB 2 QSEFTQWISRWTFSTNHRDITGLTVLIFGAFSGVLGCWSMLRMLADP----- 50
 QY 71 LIADASQTCANGHLAMNVTYHGLIMFVGIIPALFEGFGNYLMPLOIGADPMAPRRN 130
 DB 51 ----SNHLLGNHQIYLVLTIAHAFILFPMVPMVIGFGNWLVIIMIGSPDMAFPRLN 106

QY 131 NLSFWLFIAGTAMGASLFPAGGDDGOLGSGVGWLVPLST--REAGYSMDLAIFAVHUS 188
 DB 107 NISFWLIPSLICLLMSALV-----EVGVTGWTVPPLSSIQSHSGAVDLAIFSLHIS 161
 QY 189 GASSIMGAINMTITFLNMRAPGNTLHKVPLFSWSIFITAMLILALPVLAGATMLLTDR 248
 DB 162 GASSILGAVNFISTILNMRSPGQSMVRIFLFWMSILVTFALLAVPVLAGATMLLTDR 221
 QY 249 NGCTTFENRAGGDDPLLYOHILMFSGHPEVYIILDFGILSHVSTFESKRPVGLPMW 308
 DB 222 NNTSFEDASGGDDPLLYOHLFWFPGHPEVYIILDFGILSHVSTFESKRPVGLPMW 281
 QY 309 YAMVAIGVGFVYVAHMYTVGMSLTQOQSFMLATVIAVPTGIKLFESNLATWKGSSVE 368
 DB 282 YAMVSIQVIGLTVMAHMYTVGMDVTRAYFTRATVIAVPTGIKLFESNLATWKGSSVE 341
 QY 369 KSPMLAFSGFMELFTVGVGTGLVLAQGLDRAYHDTYVVAHMYVSLGATFAIFAGTY 428
 DB 342 KTPMLFAIFIEFLTIGLGLVLAQGLDISLHDYVVAHMYVSLGAVFAIFAGTY 401
 QY 429 FYMPKFSGRAPFPMMAKLHPTFFIGANTFFPQHFILGQGMRRITDYPFAALNNKVS 488
 DB 402 YWFGKITGLQYPTLGLQHFWSFTFIGNLTFMPHFLGLAGMPRLDPDVAAGNLLA 461
 QY 489 SYGAFIAPASFLEFVIFYVYTLVAGRRTRPNPWCSEF-----ADTLFWTLPSPPAHT 541
 DB 462 SYGSIYALFSTLEFFFIYVSLTNNPCIN-FPW-EFNKSKTYGVSTLEMIYISPAVYHT 519
 QY 542 FETLP 546
 DB 520 FEEMP 524

Search completed: December 22, 2001, 11:08:07
 Job time: 336 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2001, 11:01:51 ; Search time 37.53 Seconds
(without alignments)
2170.895 Million cell updates/sec

Title: US-09-712-768-2
Perfect score: 3000
Sequence: 1 MADAAIHGDHHEKGFTR.....PAHTFETLPKRSMDKPSH 557

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: SPREMBL_17:*
 - 2: sp_archaea:*
 - 3: sp_bacteria:*
 - 4: sp_fungi:*
 - 5: sp_human:*
 - 6: sp_invertebrate:*
 - 7: sp_minc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1847.5	61.6	552	2	09A300
2	1838	61.3	539	2	059631
3	1773	59.1	523	8	036520
4	1756.5	58.5	525	8	097CA6
5	1751.5	58.4	528	8	063228
6	1745	58.2	515	8	037616
7	1730.5	57.7	632	8	035812
8	1720.5	57.4	532	8	0968T5
9	1707	56.9	537	8	099988
10	1706.5	56.9	531	8	092Y18
11	1698.5	56.6	531	8	021293
12	1682.5	56.1	527	8	034882
13	1670	55.7	533	8	092Z05
14	1659.5	55.3	528	8	P92800
15	1652.5	55.1	527	8	09B0H9
16	1646	54.9	529	8	09MGB2
17	1645.5	54.9	503	8	09G4C3
18	1642	54.7	531	8	099620
19	1630.5	54.4	527	8	063225

20	1624.5	54.1	471	8	09MH04	09mh04 geranium hl
21	1621.5	54.0	469	8	099289	099289 gnetum leyb
22	1621.5	54.0	565	8	09MFM0	09mfm0 aspergillus
23	1611.5	53.7	471	8	09B746	09b746 canella win
24	1609	53.6	468	8	047441	047441 ginkgo bil
25	1608	53.6	492	8	09T255	09t255 phytophthor
26	1608	53.6	527	8	047563	047563 agrocypae ae
27	1606.5	53.5	471	8	099296	099296 phyllolcladu
28	1606.5	53.5	471	8	09MH10	09mh10 buxus sp. q
29	1606.5	53.5	471	8	09MH05	09mh05 hyseochari
30	1605.5	53.5	471	8	09B747	09b747 calycanthus
31	1601.5	53.4	471	8	09MH17	09mh17 calycanthus
32	1600.5	53.4	471	8	099306	099306 weliwitschia
33	1595.5	53.2	528	8	063229	063229 sabal palme
34	1587	52.9	528	8	09G820	09g820 acorus cala
35	1586.5	52.9	471	8	09MH28	09mh28 nelumbo lut
36	1586.5	52.9	471	8	09B736	09b736 chamaerops
37	1584.5	52.8	527	8	063226	063226 caferetia r
38	1581.5	52.7	517	8	09TAK5	09tak5 caferetia r
39	1581.5	52.7	517	8	035289	035289 asperster re
40	1580.5	52.7	517	8	035430	035430 patirlella
41	1577.5	52.6	566	8	09GAT2	09gat2 aspergillus
42	1575.5	52.5	566	8	09GAS6	09gas6 aspergillus
43	1575	52.5	535	8	09MD20	09md20 scenedesmus
44	1572	52.4	446	8	079371	079371 malpighia g
45	1571.5	52.4	517	8	035616	035616 patirlella

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	552 AA.
09A300	AC	09A300:		
ID	09A300:			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CYTOCHROME C OXIDASE, SUBUNIT I.			
GN	CC3406.			
OS	Caulobacter crescentus.			
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;			
OC	Caulobacter.			
OX	NCBI_TaxID=69394;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21173698; PubMed=11259647;			
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,			
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,			
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,			
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,			
RA	Kolonay J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,			
RA	Uttelback T., Tyrn K., Wolf A., Yarnaliev J., Ermolaeva M., White O.,			
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;			
RT	"Complete genome sequence of Caulobacter crescentus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).			
DR	EMBL; AE006001; AAK25368.1;			
DR	TIGR; CC3406;			
KW	Complete proteome.			
SEQ	SEQUENCE 552 AA; 61101 MW; E82191E54206AAC5 CRC64;			

Query Match 61.6%; Score 1847.5; DB 2; Length 552;
Best Local Similarity 61.6%; Pred No. 14e-93;
Matches 341; Conservative 70; Mismatches 136; Indels 7; Gaps 4;

QY	2	ADAAIHGDHHEKGFTRFEMSTNKDGLLYVAAGVGFISVLYMRLEMDPGV 61
DB	5	ADTHDAHDDDHKPPFAWFFSTNHKIDIGTLYILFAIAGVIGVALSLIMELMARP 64
QY	62	QYVCEGARGALIDASQCTANGHIMVWVITYGILMFEVGPALFGGNGATLMPLOGA 121
DB	65	QVFSDTGML---AOMGIFKRGKGYNAVVTAAHALIMFFVMPAMIGCGNWFVIMIGA 120

Qy	122	PDMAFPNNNNISFPLFAGTAMCVASLFAFGGCGOLSGVGWVLYPLPSF-REAGYSMDL	180
Db	121	PDMAFPNNNNISFPLLLAAMIILVLMSFVDGGRGH-GFGGGMIIYRPLSTIGHKGRAMD	179
Qy	181	ALFAVHLGASSINGAINMTTFELNMRAPOMTLHKVPLEWSFSFTTAMVLLALPVLAA	240
Db	180	ALSLHLGASSILGAINFTTTEINMRAPOMTLHRMPLFWVSVLLAFALLLSLPLAA	239
Qy	241	ITMLLTDRNCTGFENPAGGDPFLYOHILIMFEGHREPVYIIILPGGIIISHVYFSFSKP	300
Db	240	ITMLLTDRNCTGFEDPAGGDPVRFQHLWFGHREPVYIILLPGGIIISHVYFSFSKP	299
Qy	301	VEGYLPNTAVAVAGVLGVFWMAHHMTVGNLSITQOQSYEMALATVIAVPTGIKFSMIAT	360
Db	300	VEGYLMAVAVAVAGVGFVMAHHMTVGNLSINLRAYFAAATMIIVAPRGVIKFSMIAT	359
Qy	361	MWGSVERKSMMLAFCGMFLFTYGVGTGYLVLAQAGIDRAVYHDTYVVAHFHVMISLGI	420
Db	360	MWGSLSKRTKMLMAIGITFLETVGVGTGVVLSNAG-IDYLLHDTYVVAHFHVMISLGA	419
Qy	421	FAIFAGITFYMPKFSGRFAPEMAKLFHFMTFEGANVTFPQHFILGROGAPRRYIDYPEA	480
Db	420	FAIFAGITFYWEKMGVAYNEFLDAHFHMFEGVNLIFPQHFILGQGRPRRYIDYPA	479
Qy	481	FALMNKSSYCAFLAFASFLFYIYVYTLVAGRGRETRPNPWGEFADTLTWTLSPPPAH	540
Db	480	FTLMNVYSTVGVYMITVVGVFVLMVLLEALIR-BRKAEDANWCGATVTLTWTLSSPPFH	538
Qy	541	TFEFLPKRSDKH	554
Db	539	QPSRPVIAKGDHH	552

Q59631	2	PRELIMINARY:	PRT:	539 AA.
AC	059631:			
AD	01-NOV-1996 (TREMblrel. 01, Created)			
DT	01-NOV-1996 (TREMblrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)			
DE	CYTOCHROME-C OXIDASE SUBUNIT 1 (EC 1.9.3.1) (CYTOCHROME OXIDASE)			
DE	(CYTOCHROME A(3)) (CYTOCHROME AA(3)).			
GN	COXB.			
OS	Nitrobacter winogradskyi (Nitrobacter agilis).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Bradyrhizobium group; Nitrobacteraceae; Nitrobacter.			
OX	NCBI_TaxID=913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 14123;			
RA	Berben G.;			
RL	Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME			
CC	C + 2 H(2)O.			
CC	-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4			
CC	FERRICYTOCHROME C.			
CC	-1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).			
CC	-1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY			
DR	EMBL; X89566; CAA61744.1; -.			
DR	HSSP; P98002; IARI.			
DR	InterPro; IPR000883; COX1.			
DR	Pfam; PF00115; COX1; 1.			
DR	PRINTS; PR01165; CYCOXIDASE1.			
DR	PROSITE; PS00077; COX1; 1.			
KM	Copper: Heme: Oxidoreductase: Respiratory chain: Transmembrane.			
SEQ	SEQUENCE 539 AA: 59145 MW; 2639148DACA23628 CRC64;			

Query Match	61.3%;	Score 1838;	DB 2;	Length 539;
Best Local Similarity	61.8%;	Pred. No. 4.6e-93;		

Matches 345; Conservative 74; Mismatches 107; Indels 32; Gaps 7;

Qy	1	MADDAINGHON-----HEKOFEPFRMMSJNHNHDIGLTLVLAAGVGEISVLEIYVMK	53
Dp	1	MATASANDHNDHGDONHNPBG-WKRFAYSTNHKDGIMVLVAIVAGIIIGGLMSGIR	59
Qy	54	LELMDRCVOYMLCEGARLADDSQJCTANGHLMNVVTHNGLIMMEFVGIPALFGGSGNY	113
Dp	60	AELMYPGIO-----VEDPA-----HUYENFATNHGIMLIFVFWMPMAGSGGNW	103
Qy	114	LMPLQIOTARPMARPRNNJNSEFMLCTAGTMGVAASIAPAGCGDQJGSGVWLYPRLSTR-	172
Dp	104	FVRPLMOTARPMARPRNNJNSEFWLLPRAFALLISSEMGEBRGATGWTLLYAPRLSG.	163
Qy	173	EAGYSMDLAIFFAHLSSGASSIMGAIMITPLNMRAPGTLHKYPLFWSMIFITAMLIL	232
Dp	164	HGPAVNDPSTLSIHLGASSIILGAINFITTEINMRAPGTHKMYPLFWSIILVWVPLLL	223
Qy	233	ALPYLAGATMLLTDNRNFGTFRNPAGGDRPLLXOHLMEFSGHREVVYIILPRGCIISHV	292
Dp	224	ALPYLAGATMLLTDNRNFGTFRNPAGGDRPLFOHLEFFGHREPVYIILPRGCMISHV	283
Qy	223	VSTSSKRPVSGULRYUVAWATGVLRRVVAHNMHTUGSLQOQSYPLATMVLAVPRGI	352
Dp	284	ISTSSKRPLESGUAMALAMAYATIGIGRVVAHNMHTUGVSSAOQYFALATMVLAVPRGV	343
Qy	353	KIFEMITMMGGSVEFKSPYMLMAFGEMFLTEVGGVGIYLVLAQGLDRAYHDTYUUVANH	412
Dp	344	KVFSMIMTMGGSIEFKTPMLFALIGLFLTEVGGVGVVLANAGVDVMOEYUUVIANH	403
Qy	413	YVNSIGLAFALFAGIYUVMKESGRAPREMAKLNHWFFTCIANTFFPRQFLRGCKNPR	472
Dp	404	YVMGIAAFALFASCMYUWFKPMGUYMNEFIGLHWVFICIANLILFFRQHLGIQCKMR	463
Qy	473	RYIYRPAFAALMNWSSYSGAFALFASPLRFIVFYUUVL---AGRETRPMPMGCFADTL	529
Dp	464	KRVYDPDAFAGWNEISISYGAFTTG---RGVILFLGLVDARAKQOQADNPMBEGATTL	519
Qy	530	EWTLSPRPAHTEFTLPK	547
Dp	520	EWTLSSPRPFHQFSTLPK	537

```

036520 RESULT 3
ID 036520 PRELIMINARY; PRT; 523 AA.
DC 036520.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COXI OR COXI.
OS Platyomonas subcordiformis.
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Platyomonas.
OX NCBI_TaxID=3161;
RN [1]
RP SEQUENCE FROM N.A.
RA Kessler U.;
RL Thesis (1995), Institute for Plant Physiology,
  Justus Liebig University, Giessen, Germany.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9614517; PubMed=8555450;
RA Kessler U.; Zetsche K.;
RT "Physical map and gene organization of the mitochondrial genome from
  the unicellular green alga Platyomonas (Tetraselmis) subcordiformis
  (Prasinophyceae).";
RL Plant Mol. Biol. 29:1081-1086(1995).
CC -! FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
  CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1
  3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
  CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN

```

CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL: Z47795; CAA87753.1; .
DR HSP: P98002; IRL1.
DR Mendel: 21543; Plasu; cox1; 21543.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PRO1165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
SQ SEQUENCE 523 AA; 57669 MW; 3E5B17B6E749085C CRC64;

Query Match 59.1%; Score 1773; DB 8; Length 523;
Best Local Similarity 60.1%; Pred. No. 1.6e-89;
Matches 322; Conservative 82; Mismatches 104; Indels 28; Gaps 6;

QY 17 FETRMFSTNKHDKIGLLYLVAAGVGFISVLFYVRLMLDMPGVQYMCLEGARLIADAS 76
DB 4 FAQWMLSTNKHDKIGLLYLVAAGVGFISVLFYVRLMLDMPGVQYMCLEGARLIADAS 48
QY 77 QTCANGHLNVAWVYHGLIMFVFGIPALFEGFGNYLMPLOIGADMAPRPMNLSEWL 136
DB 49 QVLGNHOLYVNVITTAHAFLMFFVMPBALGEGFNWFMIGADMAPRPMNLSEWL 108
QY 137 FIAGTAVASLFAFGDGOIGSGVGVNLYPLRST--REAGYSMDLIPAVHLISGASSIM 194
DB 109 LPPSLLLSSALV-----ELGAGTGMVYPPPLSSIASHSASVDLAIFSLHSIGASSIL 163
QY 195 GAIMNTTFLNMRAPGMLLHKVPLFSWISFTAWLILALPVLAGATMLLDRNFSTTE 254
DB 164 GALNFITTLNMRGPMTHRLPLFVMSVLTATALLSLPVLGATITMLLDRNFSTTE 223
QY 255 FNPAGGDPILYOHILMFEFGHEPVYIIILPGFGLISHVSTFSKRPVGYLPMYANVAI 314
DB 224 FDPAGGDPILYOHILMFEFGHEPVYIIILPAFGILSHVSTFSKRPVGYLPMYANVAI 283
QY 315 GVLGFVVAHMHYVGMSTLQOVSFMLATVYIAVPTGKIFSMATMGGSVERKSPMLW 374
DB 284 GILGFIVMAHMHYVGLDIDITRAITFAATMIAVPTGKIFSMATMGGSVERKSPMLF 343
QY 375 AFGMFLFTVGGVTVGLVLAQAGLDRAVHDTYVVVAHFHYVMSLGAIFAFAGIYFMPKF 434
DB 344 AIGFLFLFTVGGVTVGLVLAQAGLDRAVHDTYVVVAHFHYVMSLGAIFAFAGIYFMPKF 403
QY 435 SGRAPFPMMAKLHWTFTTCANTFFPOHPLRGOMPRRYITDPEAFALNNKVSYSYGL 494
DB 404 CGLOYSSEFLQIHFMFLFPGVNIITFFPMHFLGISGMRRIPDYDPAAGNNAVCSYGYL 463
QY 495 AFASFLEFVIFVYTLVAGRRETRPNF---GEFADTLTMTLPSPPAPTEFLP 546
DB 464 SVLGALFEFVY-VYDITLSGERCKPMETVPGTSA-TLEMLTPSPAPYHTFEVP 517

RESULT 4
QYTCAG PRELIMINARY; PRT: 525 AA.
AC QYTCAG;
DT 01-MAY-2000 (TREMBLREL. 13, Created)
DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Nephroselmis olivacea.
OC Mitochondrion.

OC Eukaryota: Viridiplantae; Chlorophyta; Prasinophyceae;
OC Chlorodendrolales; Chlorodendraceae; Nephroselmis.
OX NCB1_TaxID=31312;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIES-484;
RX MEDLINE-99418884; PubMed-10488238;
RA Turmel M., Lemieux C., Burger G., Lang B.F., Otis C., Plante I.,
RA Gray M.W.;
RT "The complete mitochondrial DNA sequences of Nephroselmis olivacea and
RT Pedionomonas minor: two radically different evolutionary patterns
RT within the green algae."
RL Plant Cell 11:1717-1729(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NIES-484;
RA Turmel M., Lemieux C., Burger G., Lang B.F., Otis C., Plante I.,
RA Gray M.W.;
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF110138; AAF03191.1; .
DR HSP: P18401; IFTT.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PRO1165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
SQ SEQUENCE 525 AA; 58099 MW; 04293667EBAD4AA CRC64;

Query Match 58.5%; Score 1756.5; DB 8; Length 525;
Best Local Similarity 58.2%; Pred. No. 1.2e-88;
Matches 312; Conservative 90; Mismatches 107; Indels 27; Gaps 5;

QY 17 FETRMFSTNKHDKIGLLYLVAAGVGFISVLFYVRLMLDMPGVQYMCLEGARLIADAS 76
DB 4 FVQRMFSTNKHDKIGLLYLVAAGVGFISVLFYVRLMLDMPGVQYMCLEGARLIADAS 48
QY 77 QTCANGHLNVAWVYHGLIMFVFGIPALFEGFGNYLMPLOIGADMAPRPMNLSEWL 136
DB 49 QILAGHOLYVNVITTAHAFLMFFVMPBALGEGFNWFMIGADMAPRPMNLSEWL 108
QY 137 FIAGTAVASLFAFGDGOIGSGVGVNLYPLRST--AGYSMDLIPAVHLISGASSIM 194
DB 109 LPPSLLLSSALV-----EVGAGTGMVYPPPLSSIASHSASVDLAIFSLHSIGASSIL 163
QY 195 GAIMNTTFLNMRAPGMLLHKVPLFSWISFTAWLILALPVLAGATMLLDRNFSTTE 254
DB 164 GAINFTITFLNMRGPMTHRLPLFVMSVLTATALLSLPVLGATITMLLDRNFSTTE 223
QY 255 FNPAGGDPILYOHILMFEFGHEPVYIIILPGFGLISHVSTFSKRPVGYLPMYANVAI 314
DB 224 FDPAGGDPILYOHILMFEFGHEPVYIIILPAFGILSHVSTFSKRPVGYLPMYANVAI 283
QY 315 GVLGFVVAHMHYVGMSTLQOVSFMLATVYIAVPTGKIFSMATMGGSVERKSPMLW 374
DB 284 GILGFIVMAHMHYVGLDIDITRAITFAATMIAVPTGKIFSMATMGGSVERKSPMLF 343
QY 375 AFGMFLFTVGGVTVGLVLAQAGLDRAVHDTYVVVAHFHYVMSLGAIFAFAGIYFMPKF 434

Db 344 AVGFILFTTGFTGILLNSGDLIALHDYVVVGHFHVLSMGAVFGMFAFYWIGKI 403

Qy 435 SGAPAPENMAKHEWTFEFGANVTEPPQHELRQGMREYIDYPEAFALMNKVSYGARL 494

Db 404 TGLQYETLGLQHFWLFTFGVNTPEPMHFLGLAGMPRIPIPDYVAGWNAIASGSL 463

Qy 495 AFASFEEFVIEVYTLVAGRRETRPNMG----EPADLEWTLPSPPHPTETPL 546

Db 464 SILGALFFEVV-VYATLTGNKGNPNMADKTRDASTLEMMVSGSPARHTFHQIP 518

RESULT 5

063228 PRELIMINARY: PRT: 528 AA.

AC 063228:

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).

GN COX1.

OS Populus tremuloides (Quaking aspen).

OC Mitochondrion.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids I; Malpighiales; Saliaceae; Populus.

OX NCBI_TaxID=3693;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LEAF;

RC MEDLINE=97303197; PubMed=9159140;

RA Laroche J., Li P., Maggia L., Bousquet J.;

RT Molecular evolution of angiosperm mitochondrial introns and exons.";

Proc. Natl. Acad. Sci. U.S.A. 94:5722-5727(1997).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: U77623; AAC98473.1; -

DR HSSP: P98002; IARL.

DR Mendel: 23876; Poplm:cox1;23876.

DR InterPro: IPR000883; COX1.

DR Pfam: PF00115; COX1; 1.

DR PROSITE: PS00077; COX1; 1.

KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane.

FT NON_TER 528 528

SEQUENCE 528 AA; 57993 MW; 46600F70CD4D5047 CRC64;

Query Match 58.4%; Score 1751.5; DB 8; Length 528;

Best Local Similarity 59.0%; Pred. No. 2.3e-88;

Matches 314; Conservative 89; Mismatches 104; Indels 25; Gaps 5;

Qy 20 RWMSTNHKDIGILLYLVAGVGFISLFTVYMRLELMDPGVQYMCLEGA RLTAASQTC 79

Db 6 RMLFSTNHKDIGILLYLVAGVGFISLFTVYMRLELMDPGVQYMCLEGA RLTAASQTC 79

Qy 80 TANGHLNVMVYTHGILMEFVGPALFGSGVYMLPLDGLADMAFPKRNNSFWLFTA 139

Db 51 GGNHQLYNNVLTAAHAFIMFPMVPMIGFGWMPVILIGADMAFPRLNINSFWLLPP 110

Qy 140 GTAMGVAFLFAPGGDGLGSGVGVWVLYPLPS--TREAGSMDLTAFAVHLSGSSIMGAI 197

Db 111 SLILLSSALV-----EVSGTGWTYVPLPSLISHSGAVDLAIFSLHLSGVSSILGSI 165

Qy 198 NMTITFLMARPGMTLHKVLPFSWISFTAWLILALPVLAGAITMLTDNRNGTFENP 257

Db 166 NPTITIFMRGPGMTMHRLPFVMSVLVATFALLSLSPVLAGAITMLTDNRNFTTFEDP 225

Qy 258 AGGSDPLLYOHILMEFGHPEVYIILDPFGIISHVSTFSKRPVEGYLPMVYANVAIGVL 317

Db 226 AGGGDPLLYQHLEFPEFGHPEVYIILDPFGIISHVSTFSKRPVEGYLPMVYANVAIGVL 285

Qy 318 GGVVAHNMKTVGMSLTDQSYEMLATVIANPQGIKIFSWATMGSGVSEKSPMLAFG 377

Db 286 GFLVANHMFVTVGLVDVTRAFATMTIAVDTGKIFSWATMGSGIDQKTFMPLAVG 345

Qy 378 FMFLTVGCVTGIYLAQGLDRAVHDYVVVVAHFVYMSGALFAIPAGIYFYMPKFSGR 437

Db 346 FIFLETGLGLGYLANGLDIALHDYVVVVAHFVYMSGAVFALPAGFYWVGKITGR 405

Qy 438 APPEMAKHEWTFEFGANVTEPPQHELRQGMREYIDYPEAFALMNKVSYGARLFA 497

Db 406 TYPEPLGKHEWTFEFGANVTEPPQHELRQGMREYIDYPEAFALMNKVSYGARLFA 465

Qy 498 SFLFEVTVVTVLAG-RRETRPNMG--EPADLEWTLPSPPHPTETPL 546

Db 466 GICCFVTVVTVLSSGNONKCAPSPWALDQNSTLEMMVSGSPARHTFHQIP 517

RESULT 6

037616 PRELIMINARY: PRT: 515 AA.

AC 037616:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).

GN COX1.

OS Prototheca wickerhamii.

OC Mitochondrion.

OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorrellales;

OC Chlorrellaceae; Prototheca.

OX NCBI_TaxID=3111;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=263-11;

RC MEDLINE=94180393; PubMed=8135522;

RA Wolff G., Plante I., Lang B.F., Kueck U.;

RT "Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii. Gene content and genome organization.";

RL J. Mol. Biol. 237:75-86(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=263-11;

RC MEDLINE=93181269; PubMed=7680126;

RA Wolff G., Burger G., Lang B., Kueck U.;

RT "Mitochondrial genes in the colourless alga Prototheca wickerhamii resemble plant genes in their exons but fungal genes in their introns.";

RL Nucleic Acids Res. 21:719-726(1993).

CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).

CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: U02970; AAD12634.1; -

DR HSSP: P98002; IARL.

DR Mendel: 21560; Protol:cox1;21560.

DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1.1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1.1.
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
 KW Respiratory chain; Transmembrane.
 SEQUENCE 515 AA; 57036 MW; 2F6D1AD87E240B CRC64;

Query Match 58.2%; Score 1745; DB 8; Length 515;
 Best Local Similarity 58.2%; Pred. No. 5.1e-88;
 Matches 310; Conservative 92; Mismatches 103; Indels 28; Gaps 6;

QY 19 TRWENSTNKHDKIGLLYLVAAGVGFISLVTVYKRLLEIMDPGVQYMCLEGARLIADASQT 78
 DB 3 TRWELSTNKHDKIGLLYLVAAGVGFISLVTVYKRLLEIMDPGVQYMCLEGARLIADASQT 47
 QY 79 CTANOHLMNWNVYTHGILMFFVGPALFGFGNYLPLDGLADMAPPRNNLSFWLFTI 138
 DB 48 LNHQHLVNLVITLHAFLMIFPMPLMGFGNWFPLILGADDMAPRLNLSFWLPL 107
 QY 139 AGTAMGVASLEFAPGGDGLSGVGWVLYPLST--REAGYSMDLAFVHLGASSIMGA 196
 DB 108 PSLILVSSALV-----EVGAGTGWYVYPLPLASTASHSGSVDLAIFSLHLGVSSTLGA 162
 QY 197 INMTTFLNMRAPGMTLHKYPLFMSISITFAMLLALPLVLAGATMLLDRNGTFEEN 256
 DB 163 INICTVFENMRAPGMSMRPLFLFVAVFETAMLLLCPLVLAGITMLLDRNNTSFED 222
 QY 257 PAGGDDPLTYOHLIMFGHREYVYIIILPGFISHVYSTFSKRVFGLPMVYAMAIVG 316
 DB 223 PAGGDDPLTYOHLIMFGHREYVYIIILPGFISHVYSTFSKRVFGLPMVYAMAIVG 282
 QY 317 LGFVMAHMYVYVGMSTQOVSFYLATVAVPTGKIFSMIADMGSGVEKSPMLMAF 376
 DB 283 LGFVMAHMYVYVGMSTQOVSFYLATVAVPTGKIFSMIADMGSGVEKSPMLMAF 342
 QY 377 GFMLFTVGVGTGYVLAQGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRFSG 436
 DB 343 GFMLFTVGVGTGYVLAQGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRFSG 402
 QY 437 RAFPENAKLHEFTFICGANTFFPOHLEFGROGPRRYIDYPEAFALMNKVSYGARLAF 496
 DB 403 LQYETLGLQHLFMFLMFGVNTFFPMFLGLAGMPRIIPDPCDGAAMVASTGSLST 462
 QY 497 ASLEFVIVYVYLVAGRRETR-NPWGE--FADTLEWTLSPPAHTEETL 545
 DB 463 TAVLEFVYVYVYLVAGRRETR-NEVCPRNPWETTPGVSPLEMMILSPPAHTEETL 513

RESULT 7
 Q35812 PRELIMINARY; PRT; 632 AA.
 AC Q35812;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
 GN COX1.
 OS Sorghum sp. (sorghum).
 OG Mitochondrion.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 OC NCBI_taxid=4559;
 OX [1]
 RN SEQUENCE OF 502-632 FROM N.A.
 RP STRAIN=MUTANT 9E;
 RC MEDLINE=87051727; PubMed=3022938;
 RA Bailey-Serres J., Hanson D.K., Fox T.D., Leaver C.J.;
 RT Mitochondrial genome rearrangement leads to extension and relocation
 of the cytochrome c oxidase subunit I gene in sorghum.,"
 RL Cell 47:567-576(1986).

RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MUTANT 9E;
 RA Hanson D.K., Bailey-Serres J., Leaver C.J.;
 RL Submitted (JUN-1987) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
 AND COPPER B (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
 CC -1- FERROCYTOCHROME C.
 CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 DR EMBL: M14454; AAA70319.1; -
 DR HSSP: P98002; 1AR1.
 DR Mendel: 2179; Sorgh; COX1; 2179.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1.1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1.1.
 KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
 KW Respiratory chain; Transmembrane.
 SEQUENCE 632 AA; 70204 MW; 85575B1A84603D80 CRC64;

Query Match 57.7%; Score 1730.5; DB 8; Length 632;
 Best Local Similarity 58.5%; Pred. No. 3.9e-87;
 Matches 311; Conservative 88; Mismatches 108; Indels 25; Gaps 5;

QY 20 RWPSTNKHDKIGLLYLVAAGVGFISLVTVYKRLLEIMDPGVQYMCLEGARLIADASQT 79
 DB 6 RWPSTNKHDKIGLLYLVAAGVGFISLVTVYKRLLEIMDPGVQYMCLEGARLIADASQT 50
 QY 80 TANGHLMNWNVYTHGILMFFVGPALFGFGNYLPLDGLADMAPPRNNLSFWLFTI 139
 DB 51 TANGHLMNWNVYTHGILMFFVGPALFGFGNYLPLDGLADMAPPRNNLSFWLFTI 110
 QY 140 GTAMGVASLEFAPGGDGLSGVGWVLYPLST--REAGYSMDLAFVHLGASSIMGA 197
 DB 111 GTAMGVASLEFAPGGDGLSGVGWVLYPLST--REAGYSMDLAFVHLGASSIMGA 165
 QY 198 NMITTLNMRAPGMTLHKYPLFMSISITFAMLLALPLVLAGATMLLDRNNGTFEEN 257
 DB 166 NMITTLNMRAPGMTLHKYPLFMSISITFAMLLALPLVLAGATMLLDRNNGTFEEN 225
 QY 258 AGGDDPLTYOHLIMFGHREYVYIIILPGFISHVYSTFSKRVFGLPMVYAMAIVG 317
 DB 226 AGGDDPLTYOHLIMFGHREYVYIIILPGFISHVYSTFSKRVFGLPMVYAMAIVG 285
 QY 318 GFVMAHMYVYVGMSTQOVSFYLATVAVPTGKIFSMIADMGSGVEKSPMLMAF 377
 DB 286 GFVMAHMYVYVGMSTQOVSFYLATVAVPTGKIFSMIADMGSGVEKSPMLMAF 345
 QY 378 GFMLFTVGVGTGYVLAQGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRFSG 437
 DB 346 GFMLFTVGVGTGYVLAQGLDRAVHDTYVVAHFHYVMSLGAIFALFAGIYFMRFSG 405
 QY 438 RAFPENAKLHEFTFICGANTFFPOHLEFGROGPRRYIDYPEAFALMNKVSYGARLAF 497
 DB 406 RAFPENAKLHEFTFICGANTFFPOHLEFGROGPRRYIDYPEAFALMNKVSYGARLAF 465
 QY 498 SLEFVIVYVYLVAGR-RETRNPWG--FADTLEWTLSPPAHTEETL 546
 DB 466 SLEFVIVYVYLVAGR-RETRNPWG--FADTLEWTLSPPAHTEETL 517

RESULT 8
 Q36875

ID 096875 PRELIMINARY; PRT; 532 AA.
AC 096875:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Rhodomonas salina.
OC Mitochondrion.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Rhodomonas.
OX NCBI_TaxID=529370;
RN [1]
RP SEQUENCE FROM N.A.
RA Burger G., Lang B.F., Maier U.G., McPadden G.I., Gray M.W.;
RT "Algae with secondary chloroplasts have mitochondria that originate from the host."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC -1- COPACITOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF288090; AAG17762.1; -.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
KW SEQUENCE 532 AA; 59152 MW; 64E39A3CD29CEFE4 CRC64;

Query Match 57.4%; Score 1720.5; DB 8; Length 532;
Best Local Similarity 57.5%; Pred. No. 1.2e-86;
Matches 311; Conservative 90; Mismatches 107; Indels 33; Gaps 6;

17 FFTRMSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADAS 76
1 FTRMSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADAS 76
3 FTRMSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADAS 76
77 QCTANGHLMNVVYTHGLMFEFGVIRALRGFGNYLMDLQIGADNAFFRMNNLSFWL 136
1 QCTANGHLMNVVYTHGLMFEFGVIRALRGFGNYLMDLQIGADNAFFRMNNLSFWL 136
51 ---GGNHQLYNVLVIGHAFIMFEFVMPALIGYGNELVPMIGADVAFPRMNVSEFWL 107
1 ---GGNHQLYNVLVIGHAFIMFEFVMPALIGYGNELVPMIGADVAFPRMNVSEFWL 107
137 FLAGTMGVASLEFAPGGDGLSGVGVNLYPLSTREA--GSMDLAIFAVHLSGASSTM 194
1 FLAGTMGVASLEFAPGGDGLSGVGVNLYPLSTREA--GSMDLAIFAVHLSGASSTM 194
108 LPPALLLISSTLTGEG----GAGTGWYPPPLSSVEGSPSAIDLDGIFSLHVAAGSSTL 162
1 LPPALLLISSTLTGEG----GAGTGWYPPPLSSVEGSPSAIDLDGIFSLHVAAGSSTL 162
195 GAINMTITELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLAGATMILLDRNNGTTE 254
1 GAINMTITELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLAGATMILLDRNNGTTE 254
163 GAINMTITELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLAGATMILLDRNNGTTE 222
1 GAINMTITELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLAGATMILLDRNNGTTE 222
255 FNPAGGDDPLVQHLIMFGHREPVYIIIPRGISIHVVSTSSKRVFEGYLMVAVMAV 314
1 FNPAGGDDPLVQHLIMFGHREPVYIIIPRGISIHVVSTSSKRVFEGYLMVAVMAV 314
223 FNPAGGDDPLVQHLIMFGHREPVYIIIPRGISIHVVSTSSKRVFEGYLMVAVMAV 282
1 FNPAGGDDPLVQHLIMFGHREPVYIIIPRGISIHVVSTSSKRVFEGYLMVAVMAV 282
315 GVGFGVMAHMYTVGMSLTQOQSYFMTATVAVPTGKIFSMIAIMMGSGVEFSFPM 374
1 GVGFGVMAHMYTVGMSLTQOQSYFMTATVAVPTGKIFSMIAIMMGSGVEFSFPM 374
283 GLGFGVMAHMYTVGMSLTQOQSYFMTATVAVPTGKIFSMIAIMMGSGVEFSFPM 342
1 GLGFGVMAHMYTVGMSLTQOQSYFMTATVAVPTGKIFSMIAIMMGSGVEFSFPM 342
375 AFGEFNLFTVGVGTIVLAQAGLDRAVHDTYVVAVAHFVHVSIGALFALFAGIYMPKE 434
1 AFGEFNLFTVGVGTIVLAQAGLDRAVHDTYVVAVAHFVHVSIGALFALFAGIYMPKE 434
343 AVGEFNLFTVGVGTIVLAQAGLDRAVHDTYVVAVAHFVHVSIGALFALFAGIYMPKE 402
1 AVGEFNLFTVGVGTIVLAQAGLDRAVHDTYVVAVAHFVHVSIGALFALFAGIYMPKE 402
435 SGRAPEMAAKLHFWTFEGANVTFFQHFGLGQGNPKRXYIDYPEAFALMKNVSSIGA-F 493
1 SGRAPEMAAKLHFWTFEGANVTFFQHFGLGQGNPKRXYIDYPEAFALMKNVSSIGA-F 493

DB 403 TGFQYENLGVHFWCTFVGVNLTFFQHFGLGAGMPRRIPDYSABMNNLSGSYF 462
QY 494 LAFASLEFIVIVYTVAGARRETRPMP-----GEPAIDLEMTLSPPAHFEETL 545
1 LAFASLEFIVIVYTVAGARRETRPMP-----GEPAIDLEMTLSPPAHFEETL 545
DB 463 SVAFIFLEFVLI--YETLTNMEQCEVPMKFSMEDVKNDEYTLVMSPPAFHFEV 520
1 SVAFIFLEFVLI--YETLTNMEQCEVPMKFSMEDVKNDEYTLVMSPPAFHFEV 520
QY 546 P 546
1 P 546
DB 521 P 521

RESULT 9
ID 099988 PRELIMINARY; PRT; 537 AA.
AC 099988:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Porphyra purpurea.
OC Mitochondrion.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RA Burger G., Saint-Louis D., Gray M.W., Lang B.F.;
RT "Complete sequence of the mitochondrial DNA of the red alga, Porphyra purpurea. Inverted repeats, sequence polymorphisms, and cyanobacterial
RT Introns."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC -1- COPACITOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF114794; AAD003116.1; -.
DR HSSP: P98002; IAR1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
KW SEQUENCE 537 AA; 59944 MW; EA92FBLCTDFA59DD CRC64;

Query Match 56.9%; Score 1707; DB 8; Length 537;
Best Local Similarity 57.2%; Pred. No. 6.3e-86;
Matches 310; Conservative 89; Mismatches 101; Indels 42; Gaps 7;

20 RFWSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADASOTC 79
1 RFWSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADASOTC 79
11 RFWSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADASOTC 55
1 RFWSTNHRDGLDLYLVAGVGVFISLFTYMRLEMDPGVQYMCLEGARLIADASOTC 55
80 TANGHLMNVVYTHGLMFEFGVIRALRGFGNYLMDLQIGADNAFFRMNNLSFWL 136
1 TANGHLMNVVYTHGLMFEFGVIRALRGFGNYLMDLQIGADNAFFRMNNLSFWL 136
56 LNHQVYVWVLTETHAFILMFEFVMPALIGYGNELVPMIGADVAFPRMNVSEFWL 115
1 LNHQVYVWVLTETHAFILMFEFVMPALIGYGNELVPMIGADVAFPRMNVSEFWL 115
QY 137 ---FLAGTMGVASLEFAPGGDGLSGVGVNLYPLSTREA--GSMDLAIFAVHLSGAS 191
1 ---FLAGTMGVASLEFAPGGDGLSGVGVNLYPLSTREA--GSMDLAIFAVHLSGAS 191
DB 116 SLCLLGSAM-----VEWAGAGTWLPLPLSSIOSHSGAVALDAIFSLHLSGAS 164
1 SLCLLGSAM-----VEWAGAGTWLPLPLSSIOSHSGAVALDAIFSLHLSGAS 164
QY 192 SIMGAINMTITELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLAGATMILLDRNNG 251
1 SIMGAINMTITELNMRAPGMLTHKVPLEFSMIFITAMILLALPVLAGATMILLDRNNG 251


```

Db 165 SYLGAINETITIFNMNRGQSMYRIPLFVMSILITAFELLLAVLAGALITMLTDNEN 224
QY 252 TTFNPNAGGDPILYOHILMFEGHPEVYIIITLPGGIISHVSTFSKRPVGYLPMVYAM 311
Db 225 TTFDPSSGGGDPVLQHLFEGHPEVYIILPFGGIVSHVSTFSKRPVGYIGMIYAM 284
QY 312 VAIGLVGVVAHHMYVGSLSLTQOSYFMLATVIAVPTGIKIFSWATMGGSVEKSP 371
Db 285 LSLIILFETIWAHHMYVGLDVIDRATATMIIVPTGIKIFSWATMGGSVEKSP 344
QY 372 MLMAFGMEFLTVGVTGIVLAQGLDRAVHDYVVAHHEHYVMSLCAIPAIAGIYFYM 431
Db 345 MLFAIGFIFLFTIGLIGLILANGSLDISLHDYVVAHHEHYVMSLCAIPAIAGIYFYM 404
QY 432 PKESGAFPEMAKLFHETFEIGANTVEFPOHPLGROGMPRRYIDYEAFALNNKVSSTG 491
Db 405 EKSFGYSELGDIHWGTFEIGVNLTFEPMHFLGLAGMRRILPDYDPSVAGNNTASYG 464
QY 492 AFLAFASFLFEIVYVTVLAGRR-ETRPNDPGEFAD-----TLEWTLSPPAHTEET 544
Db 465 SYVALFSTLFEFVLYVFTLVTPRKVPARNPW-NEDSKIGSTLLEWISSPAYHTFNE 523
QY 545 LP 546
Db 524 IP 525

RESULT 10
092Y18 PRELIMINARY; PRT; 531 AA.
AC 092Y18;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Gracilariopsis lemaneiformis.
OG Mitochondrion.
OC Eukaryota; Rhodophyta; Florideophyceae; Gracilariiales; Gracilariaceae;
OC Gracilariopsis.
OC NCBL_TaxID=2782;
RN [1]
RP SEQUENCE FROM N.A.
RA Lang F.B.F., Goff L.J.;
RA Submitted (Dec-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL: A118119; AND17226.1; -.
DR HSSP: P98002; IARI.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS: PRO1165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
SQ SEQUENCE 531 AA; 59067 MW; 4C0E66950DE377B6 CRC64;

```

```

QY 19 TREMNSTNKDGLGLYLAAGVGFISVLFYMYMLEMDPGVQWCEGARLIADASOT 78
Db 10 SRWIFSTNHKIDISTLLILFAGFSGLGCGMSLIRMLAQDPGNQFL----- 55
QY 79 CYPNGGLMMVMVYVHGLIMMFPGIAPALFGFGNVLPIQIAPAPMAPRNNLSFMLEI 138
Db 56 -LGNHDIYVLLITAHAEFLIEFYVMPVMIGFGNNLVPMISSPDAFPRLNISFMLEP 114
QY 139 AGTAMVASLLEAFGGDGLSGGVWVLYPPLST--REAGYSMDLAIFAVHLSGASSINGA 196
Db 115 PSLCLLLASAIY-----EVGVGTGWTVYRPLPSIQSHSGGANDLAIFSLHISGASSILGA 169
QY 197 INMTTFELMMRAGMTLHVPLFMSISFTAMILLALDVLVLAGATMLTDNFGSTTFPN 256
Db 170 INFSTILMMNRGQSMYRIPLFVMSILITAFELLLAVLAGALITMLTDNENFTTFED 229
QY 257 PAGGDPILYOHILMFEGHPEVYIIITLPGGIISHVSTFSKRPVGYLPMVYAMVATGV 316
Db 230 PAGGDPVLQHLFEGHPEVYIILPFGGIVSHVSTFSKRPVGYIGMIYAMVSTGV 289
QY 317 LGRVVAHHMYVGSLSLTQOSYFMLATVIAVPTGIKIFSWATMGGSVEKSPMLMAF 376
Db 290 LGRVVAHHMYVGLDVIDRATATMIIVPTGIKIFSWATMGGSVEKSPMLMAF 349
QY 377 GMELEFTEGVTGIVLAQGLDRAVHDYVVAHHEHYVMSLCAIPAIAGIYFYMPKPSG 436
Db 350 GFLFELFTIGLIGLVANSGLDISLHDYVVAHHEHYVMSLCAIPAIAGIYFYMPKPSG 409
QY 437 RAPEWMAKLFHETFEIGANTVEFPOHPLGROGMPRRYIDYEAFALNNKVSSTGAPLAF 496
Db 410 VQYRELLOIHFWSFTFICVNLTFEPMHFLGLAGMRRILPDYDPSVAGNNTASYSYAL 469
QY 497 ASFLFEIVYVTVLAGRR-ETRPNDPGEFAD-----TLEWTLSPPAHTEETLP 546
Db 470 ESTLFEFVL-VFVSILTSKNPCVSAWNEQSITKGNSTLEWTVSPPAVHFEEMP 524

RESULT 11
021293 PRELIMINARY; PRT; 531 AA.
AC 021293;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN COX1.
OS Reclinomonas americana.
OG Mitochondrion.
OC Eukaryota; core jakobids; Reclinomonas.
OC NCBL_TaxID=48483;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC50394;
RX MEDLINE=97311393; PubMed=9168110;
RA Lang B.F., Burger G., O'Kelly C.J., Cedergren R., Golding G.B.,
RA Lemeux C., Sankoff D., Tumeit M., Gray M.W.;
RA "an ancestral mitochondrial DNA resembling a eubacterial genome in
RA miniature."
RA Nature 387:493-497(1997).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50394;
RA Lang B.F., Burger G.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4

```

Query Match 56.9%; Score 1706.5; DB 8; Length 531;
 Best Local Similarity 57.5%; Pred. No. 6,7e-86;
 Matches 308; Conservative 93; Mismatches 106; Indels 29; Gaps 5;

```

CC -1- FERRICYTOCHROME C.
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL: AF007261; AAD11923.1; -.
DR HSSP: P98002; IAR1.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.1.
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1.1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
SQ SEQUENCE 531 AA; 58745 MW; A6B130AFB3C1ACF CRC64;

Query Match 56.6%; Score 1698.5; DB 8; Length 531;
Best Local Similarity 57.6%; Pred. No. 1.8e-85;
Matches 311; Conservative 83; Mismatches 115; Indels 31; Gaps 5;

QY 17 FETRMFMSNRHNDIGLLVLAAGVGVISLVETVYMRLEMDPGVQVYCLGARIADAS 76
DB 5 FVKRFESFNHNDIGALLTMEFTGATITTSVMRLGLPG-----N 49
QY 77 QCTANGHLINNVYTHGILMFEVGIPALPGEGNYLMPLOIGADMAFPRMNNLSFML 136
DB 50 QILQGNHQLVNLITAHGLMLFMVYVMPILLGFCGNFVPLILGAPDMAFPRLNISFML 109
QY 137 FIAGFMGVASLEFAPGGGOLGSGVGVLYPPLS--TREAGYSMDLAIFAVHLSGASSIM 194
DB 110 LPPALLLFFSALV-----EVGAGTGMVAYPPLSGIOSHSGASVDLAIFSLHSGASSVL 164
QY 195 GAINNITTFELNARAPGMTLHKVPLFSWISFIATMILLALPAGATIMLLDRNFGTTF 254
DB 165 ASINITTTFNKRAQGMTHRMLPLFWSILVTSPLVPALEVLGAGTITMLLDRNFGTTF 224
QY 255 FNPAGGDEPILYOHLIMEFGHPEVYIIILPGFIISHVSTFSKKRVGYLPMVYAVAI 314
DB 225 FDPAGGDEPLVFOHLIMEFGHPEVYIIILPGFVSVSHVISAFSRRPIIDGMYAMSSI 284
QY 315 GVLGFGVVAHHVYTGMSLTQOSYFMLATVAVPTGIKTSWITMMGSGVEFSPMLM 374
DB 285 GVLGFGVVAHHVYTGMDVTRAYTAATMVAITPTGIKTSWITMMGSGTELKAPMLE 344
QY 375 ARGFMEFLFVGVTGIVLAOAGLDRAVHDTYVVVAHFVYVSLAIFAIFAGIYFPMKE 434
DB 345 AVGFVFLFEGGLTGVLNSGLDIALHDTYVVVAHFVYVSLMGAMPAIYAFYWFQKI 404
QY 435 SGRAPPEMAAKLHEFTFFIGANVTFFPQHLGROGMPRRIIDYPPAFALMKNVSSYGAFL 494
DB 405 TGYOYPEKLAOVQFWTTFGVNMLFEFPHFGLSGMPRIIDYPPAFSGMAVSSYGSVL 464
QY 495 AVASLEFIVIVYTVLVAGRRTRRPNMG-----EFADPLEWLTSSPPAFETFLP 546
DB 465 TTFSTIILMYIYRILTLDSVK-CGNDPMLAVGEGKEHFALENTLISPLSHTFEVP 523

RESULT 12
034882 PRELIMINARY: PRT: 527 AA.
AC 034882:
DB 01-NOV-1996 (TREMBLrel. 01, Created)
DB 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DB 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COX1 OR COX1.
OS Lycopersicon esculentum (Tomato).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asterales; eunsterids I; Solanales; Solanaceae; Solanum.
OK NCBI_TaxID=4081;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=PONDEROSA; TISSUE=LEAF;
RA Koh-Ichi K.;
RU Submitted (SEP-1990) to the EMBL/Genbank/DBD databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PONDEROSA; TISSUE=LEAF;
RX MEDLINE=96155620; PubMed=8575013;
RA Kadowaki K., Ozawa K., Kazama S., Kudo N., Akiyama T.;
RT "creation of an initiation codon by RNA editing in the cox1 transcript
RT from tomato mitochondria";
RL Curr. Genet. 28:415-422(1995).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC -1- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL: X54738; CA38535.1; -.
DR HSSP: P98002; IAR1.
DR Mendel: 2048; Lycos: COX1.2048.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1.1.
DR PROSITE: PS00077; COX1.1. UNKNOWN_1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 527 AA; 57652 MW; FFL189A42849DAFDB CRC64;

Query Match 56.1%; Score 1682.5; DB 8; Length 527;
Best Local Similarity 57.5%; Pred. No. 1.3e-84;
Matches 306; Conservative 86; Mismatches 115; Indels 25; Gaps 5;

QY 20 RWFSTNHNKIDIGLLVLAAGVGVISLVETVYMRLEMDPGVQVYCLGARIADASQNC 79
DB 6 RWFSTNHNKIDIGLLYTFFGAIAVYMGTCESVILRMLARG-----DQL 50
QY 80 TANGHLINNVYTHGILMFEVGIPALPGEGVNLMPLOIGADMAFPRMNNLSFMLFTA 139
DB 51 GGNHQLVNLITAHAFIMFFMVPAMIGGVMVSPILIGADMAFPRLNISFMLLPP 110
QY 140 GTANGVASLEFAPGGGOLGSGVGVLYPPLS--TREAGYSMDLAIFAVHLSGASSIMGAI 197
DB 111 SLLLSLSALV-----EVGSGTGMVYVPLSGITSHSGAVDLAISLHLSGVSSILGSI 165
QY 198 NMITTFNMRAPGMTLHKVPLFSWISFIATMILLALPVLGATIMLLDRNFGTTFNF 257
DB 166 NPTTIFNMRGPGMTMRSPFLFWSVLYTAFAPLLSLPVLGATIMLLDRNNTTFSDP 225
QY 258 AGGDDPLLYOHLIMEFGHPEVYIIILPGFIISHVSTFSKKRVGYLPMVYAVAI 317
DB 226 AGGDDPLLYOHLIMEFGHPEVYIIILPGSGIISHVSTFSKRVGYLPMVYAVAI 285
QY 318 GFVVAHHMNTYVGSLSLQOSYFMLATVAVPPIGKIFSMIATMMGSGVEFSPMLAPG 377
DB 286 GFLVAHHMNTYVGLDVTRAYTAATMVAITPTGIKTSWITMMGSGTELKAPMLE 345
QY 378 FMPFLVGVGVGVIVLAOAGLDRAVHDTYVVVAHFVYVSLAIFAIFAGIYFPMKFSGR 437
DB 346 FIEFLTIGLGVIVLANSGLDIALHDTYVVVAHFVYVSLMGAVPALPAGHYVWQKIFGR 405
QY 438 APPEMAAKLHEFTFFIGANVTFFPQHLGROGMPRRIIDYPPAFALMKNVSSYGAFLAFA 497
DB 406 TYPETLGLIHFWITFEFVNMTFFPHFGLSGMRRIRIDYPPDAVAGWNNALSSGYSISV 465

```

RT structure and organization of the mitochondrial genome of the
 RT unicellular red alga *Cyanidioschyzon merolae* deduced from the complete
 RT nucleotide sequence.";
 RL Nucleic Acids Res. 26:5190-5198(1998).
 CC -1 FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B (BY SIMILARITY).
 CC -1 CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
 CC FERROCYTOCHROME C.
 CC -1 COPROCTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
 CC -1 PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1 SIMILARITY: ANALOGY TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 DR EMBL: D89681; BAA34660.1; -.
 DR HSSP: P98002; IARL.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1; 1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1; 1.
 KM Copper: Heme: Inner membrane; Mitochondion; Oxidoreductase;
 KW Respiratory chain; Transmembrane.
 SQ SEQUENCE 533 AA; 59171 MW; 4006CF9B13DE62DF CRC64;

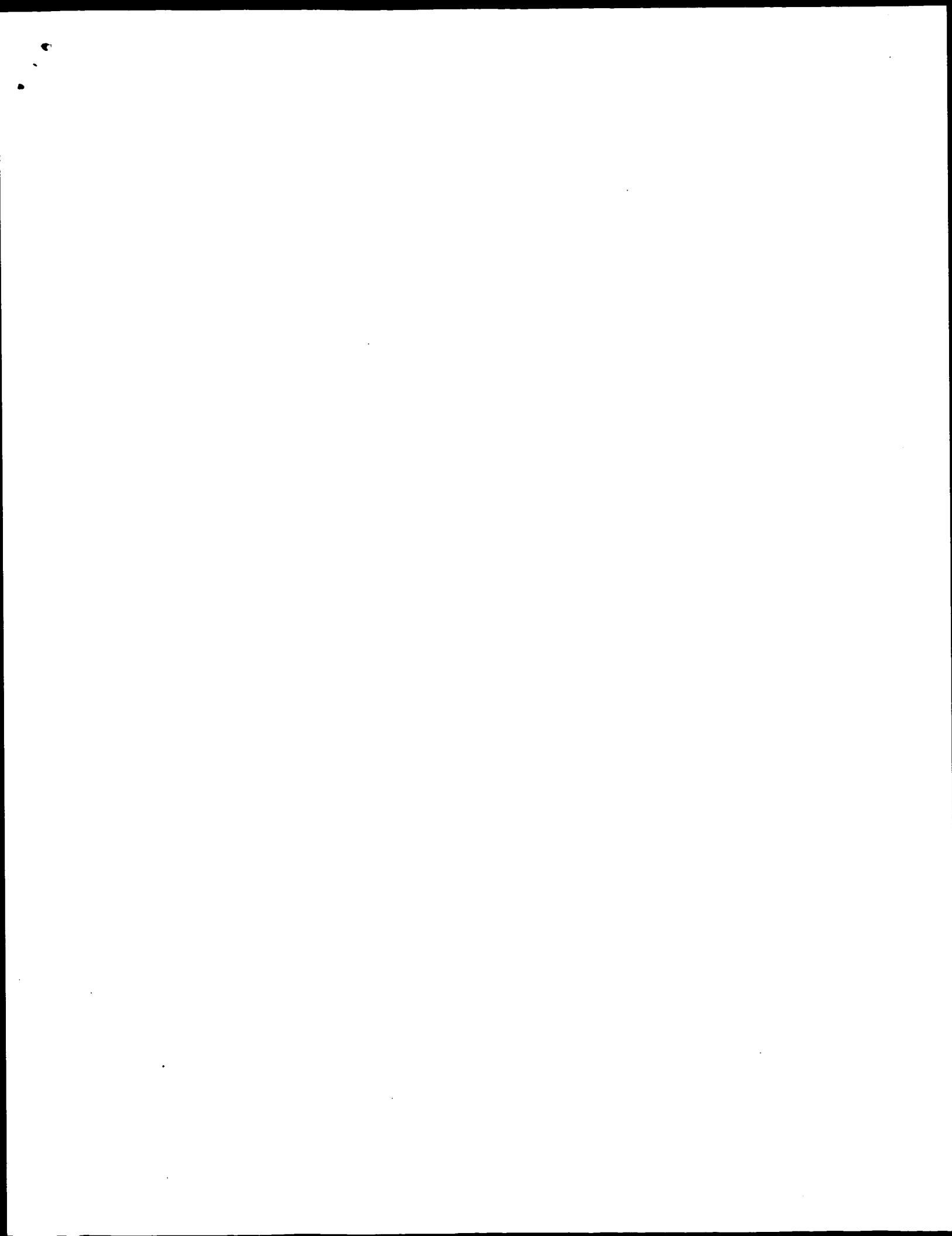
Dd	227	YDPAGGSRPLVQNHLEMFEGHEVYILLPAGFIISGVVATFESRRKPYGLGVYALISI	266
Qy	315	GVLAGVVAHHMYTQMSLTQOQSYEMLATMYIAVPTGKITISYATIMGGSVERKSPMLM	374
Dd	287	GILGLGVANHMFYGMVDVTRAFITTSIIIAVPTGKVESWTATWEGSIHLQTMPLF	346
Qy	375	AFGEMLFTYGVGTGYVLAQGLDRAYHDYVVVAHNFYVMSLGAIFAIFYEMRKE	434
Dd	347	AI GFILFTYVGILGYVLSNGSLDVAHFHDYVVVAHNFYVMSGLGIFAGFYTWGKI	406
Qy	435	SGRAPEEAAKLFHMTFFIGANMFFQHLFGROGMRRYIDYEAPALNNKYSYATFL	494
Dd	407	SGKQYPEVLGOIHFLITFLFLGNLTFEPHNFGLGIDPRIPDYDAAGMWLVATFESYV	466
Qy	495	A-FASFLFFIIFYVTVLAGRRTRPNMGEFAD-----PLEVLTSPSPRAHFEET	544
Dd	467	TLEGLTLFFIIVYV-TLLEG-QVCFEDNMWKYLDDEQTSQNSSTLLEVLVSPRAYHTIYE	524
Qy	545	LP 546	
Dd	525	IP 526	

	RESULT	14	
P92800	ID	P92800	PRELIMINARY; PTI: 528 AA.
AC	PROSITE	PS00007	
DT	01-MAY-1997	(TEMBREL_03, Created)	
DT	01-MAY-1997	(TEMBREL_03, Last sequence update)	
DT	01-JUN-2001	(TEMBREL_17, Last annotation update)	
DE	CYTOCHROME C OXIDASE POLYPEPTIDE I (NC 1.9.3.1).		
GN	COXI.		
OS	Pyallaea littoralis.		
OG	Mitochondrion.		
OC	Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;		
OC	Acinetospiraceae; Pyallaea.		
OX	NCBI_Taxid=2885;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Fontheim J., Goux D., Kloareg B., Loiseaux-de Goer S.;		
RL	J.Mol. Evol. 0:0-0(0).		
CC	- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNTS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. COI IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BINUCLEAR IRON-COPPER CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).		
CC	- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4 FERRICYTOCHROME C.		
CC	- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).		
CC	- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.		
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).		
CC	- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.		
DR	EMBL: Z72500; CAAG6581.1; -		
DR	HSSP: P98002; IARI.		
DR	Mendel: L21583; PYLI:cox1;21583.		
DR	InterPro: IPR002106; AA:Trna_Ligase_II.		
DR	InterPro: IPR006883; COX1.		
DR	Pfam: PF00115; COX1; 1.		
DR	PRINTS: PR01165; CYCOXIDASE1.		
DR	PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.		
DR	PROSITE: PS00077; COX1; 1.		
KW	Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;		
KM	Respiratory chain; Transmembrane		
SO	SEQUENCE 528 AA; 58015 MW; 11E3DDB9A055BAE CRC64;		
<hr/>			
Query Match	55.3%;	Score 1659.5;	DB 8; Length 528;
Best Local Similarity	57.5%;	Pred. No. 2,4e-83;	
Matches 308;	Conservative 85;	Mismatches 106;	Indels 37; Gaps

Thu Dec 27 08:21:36 2001

us-09-712-768-2.rpt

Page 11



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 08:09:34 ; Search time 2710.12 Seconds

(without alignments)
10190.057 Million cell updates/sec

Title: US-09-712-768-1

Perfect score: 1674

Sequence: 1 atgcagacgcgcacatca.....acaagcatccctgcactaa 1674

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb-ba:*
2: gb-hcg:*
3: gb-in:*
4: gb-om:*
5: gb-ov:*
6: gb-pat:*
7: gb-ph:*
8: gb-pl:*
9: gb-pr:*
10: gb-ro:*
11: gb-sts:*
12: gb-sy:*
13: gb-un:*
14: gb-vi:*
15: em-ba:*
16: em-fun:*
17: em-hum:*
18: em-in:*
19: em-om:*
20: em-or:*
21: em-ov:*
22: em-pat:*
23: em-ph:*
24: em-pl:*
25: em-pl:*
26: em-to:*
27: em-sts:*
28: em-sy:*
29: em-un:*
30: em-vi:*
31: em-htgo-hum:*
32: em-htgo-inv:*
33: em-htgo-rod:*
34: em-htg-inv:*
35: em-htg-rod:*
36: em-htg-other:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1674	100.0	1674	6	AX113986
2	1091.4	65.2	2629	1	PDCCYCA
3	991	59.2	3211	1	PDCCOX2
4	979.2	58.5	2166	1	RSCOXSI
5	754.4	45.1	286550	1	SME591785
6	750.2	44.8	11188	1	AE006001
7	740.4	44.2	10835	1	AE008010
8	719.6	43.0	346510	1	AP003011
9	709.2	42.4	4380	1	NMCCOXABC
10	708.8	42.3	1830	1	BJCCOXAG
11	708.8	42.3	2000	1	BJCCOXA
12	708.8	42.3	8121	1	BJA242552
13	708.8	42.3	31495	1	BJU33883
14	654	39.1	1707	1	RLCCOXA
15	515.6	30.8	12829	1	AE004449
16	454.4	27.3	1406	8	AF020565
17	454.4	27.1	1412	8	AF020564
18	450.6	26.9	1415	8	AY009436
19	450.6	26.9	4815	8	MIRSCOXI
20	449	26.8	1413	8	AF193955
21	449	26.8	166924	8	MATGENB
22	447.6	26.7	1415	8	AF020585
23	447.4	26.7	1584	8	PTU77623
24	445.8	26.6	1413	8	AF193962
25	444.2	26.5	1415	8	AY009437
26	444.2	26.5	1415	8	AY009447
27	442.6	26.4	1409	8	AF020566
28	442.4	26.4	1409	8	AF020576
29	441.4	26.4	1881	8	MTRACOI
30	441.4	26.4	2048	8	ACU46764
31	441.4	26.4	3059	8	MTRACOXI
32	441.4	26.4	1409	8	AF020572
33	440.8	26.3	1415	8	AF020584
34	440	26.3	1701	8	BVCCOXI
35	440	26.3	169799	8	AP000396S2
36	439.8	26.3	2149	8	MTRACOXI
37	439.8	26.3	2223	8	M14454
38	439.8	26.3	2345	8	SRGWTCCOXI
39	439.6	26.3	1415	8	AF020573
40	436.8	26.1	2481	8	AB024985
41	436.6	26.1	1955	8	MTRACOXI
42	434.6	26.0	1413	8	AF193968
43	434.6	26.0	16537	5	AB047553
44	433.6	25.9	1413	8	AF193967
45	428.2	25.6	1415	8	AY009454

ALIGNMENTS

RESULT	1	LOCUS	AX113986	SEQUENCE	1	from Patent	EP1103603.	PAT	01-MAY-2001
DEFINITION	AX113986	Sequence 1	1674 bp	DNA					
ACCESSION	AX113986								
VERSION	AX113986.1	GI:13940146							
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									

1.1674
/db_xref="taxon:442"

ORGANISM Paracoccus denitrificans
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Paracoccus.
REFERENCE 1 (bases 1 to 2629)
AUTHORS Raitio, M.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1990) M. Raitio, DEPT OF MEDICAL CHEMISTRY,
UNIVERSITY OF HELSINKI, SILTAVUORENENGERR 10, 00170 HELSINKI 17,
FINLAND
2 (bases 1 to 2629)
Raitio, M., Pispä, J. M., Metsä, T. and Saraste, M.
Are there isoenzymes of cytochrome c oxidase in Paracoccus
denitrificans?
JOURNAL FEBS Lett. 261 (2), 431-435 (1990)
MEDLINE 90184495
FEATURES
source 1. 2629
location/Qualifiers
/organism="Paracoccus denitrificans"
/strain="PD122"
/sub_strain="ssp. MR-3"
/db_xref="taxon:266"
/clone_lib="EMBL3 lambda"
233..2544
/gene="cyc A"
233..292
/gene="cyc A"
233..700
/gene="cyc A"
/EC_number="1.9.3.1"
/note="cytochrome-c oxidase"
/codon_start=1
/transl_table=1
/product="cytochrome c 550"
/protein_id="CA68820.1"
/db_xref="GI:45478"
/db_xref="SWISS-PROT:P00096"
/translation="MKISTYATLAITLALPAAQDDAAKGEKEFKKACAMHIOAP
DGDIDIKGKTPMLYGVGRKIASBGFYGEILVYARKNPDLTETADLLEYVD
PKPMLVMTDCKAKTKMTFKMKGNQADVAVFLAONSPDAGDEAABESN"
293..2544
/gene="cyc A"
/EC_number="1.9.3.1"
/note="cytochrome c 550"
/product="cytochrome-c oxidase"
731..765
/gene="cyc A"
871..2547
/gene="cta DIT"
871..2547
/gene="cta DIT"
/EC_number="1.9.3.1"
/note="cytochrome-c oxidase"
/codon_start=1
/transl_table=1
/product="cytochrome c oxidase beta subunit I"
/protein_id="CA68821.1"
/db_xref="GI:45479"
/db_xref="SWISS-PROT:P98002"
/translation="MADVAHGHGDHDTGPFTRMFMSTNHNKDIGILYFTAGIVGL
ISVCFYMRMLOHPEVOYMLLEGATLADASECPNGLMNWITTYGVLMFV
VITALRGKNTFRPLHIGAPDAPFRNLNLSWYVCGALVASILLAGMDOKS
GKAVNLPPLSTIEAGISMDLAFVAVHSGASSILGAINITFLNKRAGMTLFV
LFMSVFTPLMLLSLPLVLAGITMLMDNRNGTGFDPAGGDPVLYOHILTFPGH
PEVYIILPGFGIISHVSTFAKPIFGYLPNVLAMALGILGFVVAHMYAGVSL
TOQAYFLATMTIAVPTGIVKFSMATMGSGIEFKIPYMLMAFGPLFLFVGGVTV
LSQAPLDVYHDTYVAVHMYVMSLGAIVYAGVYMTGKSGSGOYPRMAQOLFM
MFTGSLIFEPHIFCRGCMRPRTYIDYPEFYWNMISSIGATISFASPLFTGIVF
YTLPAKRNVVPMNMDHDLTLETLSPPEHTFTLKRBDWRAHAH"
2586..2615
terminator 457 a 872 c 771 g 529 t
BASE COUNT
ORIGIN

Query Match 65.2%; Score 1091.4; DB 1: Length 2629;
Best Local Similarity 78.6%; Pred. No. 9.1e-219;
Matches 1318; Conservative 0; Mismatches 356; Indels 3; Gaps 1;
QY 1 atggcaagccgacccatcagcgca---tgaccacatgagaagaagccttccacg 57
Db 871 ATGCAGACGACGCGGTTCACGGCCAGGACATCATACACCGCGGCTTCCAC 930
QY 58 cgcgtggtcattgctaccacaacaaagacatgctgctataccttgcagcgtg 117
Db 931 CGGTGTTCAATGTCACAAACACACAGATATCGTATCCCTTACGCTTCCAGCGCGG 990
QY 118 gttgtgttcatctccatccgctgctacacgtcctacacgtcctgaagtcgac 177
Db 991 ATGCTGCGCCGATCTCGGTATGCTTACCGCTCTATATGGGATGGAACCTGCAGCATCCG 1050
QY 178 gttgttgaatcatgctgcttgaagcgcaagtcgtgacgagtcgacagatgt 237
Db 1051 GGCCTGCATATCATGTCCTGGAAGCGCGCGCTCTATCCGACGCGCTGCGGGAATGC 1110
QY 238 accgcaagcagacacccctgtggaagctatgcttaccacatgctatctcagatg 297
Db 1111 ACCCGAAGCAGACACCTGTGGAACGTATGATCATCACGCGGCTCATGATGTTTC 1170
QY 298 ttgtgtggtatcccgcatgttctggtgttgcgttaacatctgtagccgtgcaatc 357
Db 1171 TTGCTGCTGATCCCGGACATGCTTGGGGTTTGGGACACTATTTCATGCGGCTGCAATC 1230
QY 358 ggcgtccgagatagccttcccgatagaacaacctgcttgcgtgttcatatgac 417
Db 1231 GGCCTCCCGACATGAGCTTCCCGCGGCTGAACACCTCTCTAGTAGATGTGCTGTC 1290
QY 418 ggtacccgagtgagcgtgctgctgttctgcaaccggcgagtgacgtcagcgtg 477
Db 1291 GGCCTGACCTCGGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1350
QY 478 ggcgt 537
Db 1351 GCGCTGCGGTGCGGTGCTTACCCGCGCTGTGACACGACGCGGCGGTATTCATGAGAC 1410
QY 538 ctgcgatttctcggtgtacattgtgctgctcctcgtacatctgagcgagatca 597
Db 1411 CTGCGATCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1470
QY 598 atcagacaccttgaacatgagcgcccgagatgacgtgacaaagtgccgtgttc 657
Db 1471 ATCACACCTTCTCTCAACATGCGCGACCGGGATGACGCTGTCAAGGTGCGCTGTTG 1530
QY 658 tctgtgctgactcttatacagcgtctgctatcctctgtgagcgtgctgctgctgt 717
Db 1531 GCGTGTGCGGTCTTCAACACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1590
QY 718 gcaatcacatgctgtgacagacgttaactctgcaagacaccttcaactcgtgac 777
Db 1591 GCGATCACCATGTGCTGATGAGACCGCAACCTTGGGACGAGTTCATCCGCGCGCG 1650
QY 778 ggcgtgagacggaattctgtaacacacacatctgtgttcttctggaaccggaagtgc 837
Db 1651 GCGCGGACACCGGCTGCTTACACGACATCTGTGTTCTTCCGCGCATCCGAGGCTCTAT 1710
QY 838 atcaatcttgcgcgcttggacatcaatcagcagtcgtgacaccttccgaagaag 897
Db 1711 ATCATCATCTGCGCGGCTTGGCATCATCAGCAGCATGCTTCGACCTTCCGCAAGAG 1770
QY 898 cgcgtcttggttaactctgcagatgctctatgaaatggtgcaatcgtgttctcggcct 957
Db 1771 CCGATCTTCCGCTACCTGCGGATGCTGCGCATGCGCGGCGGATCCCTGCGGCTTC 1830
QY 958 gtcgtctgagcaccacatgataacggtgtgctgtgctgacagcaatctactac 1017
Db 1831 GTCGTGCGGCGACCATGATGATACGCGCGGATGCTGCTGACCCAGAGGCTATTTTC 1890
QY 1018 atgctgtcacaacatgtagtcggtgtgcgacccgcatlaagatcttctgtgacgc 1077

Db	1891	ATGCTGGCGACCAATGACCAATCGCGTGTCCACCGCGATCAAGGATCTTTCGTGTGATCGGG	1950
Qy	1078	acgaatctggaggcgagctcgatgagttcaaatcgcgcgatctcgtggcctttggcttatcg	1137
Db	1951	ACCAATGTGGGGCGCGCACATCGAATTCACAAACGCCGATGCTGTGGGCGCTTCGGCTTCGG	2010
Qy	1138	ttctctgttaccagtgaggctgt	1197
Db	2011	TTCCCTGTCACCGTCGCGCGCGGTGTACCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2070
Qy	1198	gcataatcagcagacacttaacgt	1257
Db	2071	GTCATATCGACACACTTATAGT	2130
Qy	1258	atctttgcgactcttcgcgcggtatctacttcttacctgtccgaagttcttcggtcgcgcttc	1317
Db	2131	GTGTTCCGCACTTCCTGCGCGCGGTCTATTACTGTGATCGCGCAAGATGTGCGCGCGCAATAC	2190
Qy	1318	ccggaatggagctgtcaaaagctcagctcttgagactcttcaatctcgtgtgcagcagcttc	1377
Db	2191	CCGGAATGGCGGGCGCGACGCTGCATTTCTGTGATATGTTTCAATCGCGCTCGAATCGATCTTC	2250
Qy	1378	ttcccgcaagaactccctcgt	1437
Db	2251	TTCCCGCGAGCACTTCCTGTGGCGCGCGAGGATGCGCGCGCGCTATATGCACTATCGGTC	2310
Qy	1438	gctcttcgagctgt	1497
Db	2311	GAGTTCCGCTTATGTGACMACATCTGCTGCATTCGCGCGCTATATCTCTCTCGGTCCTTC	2370
Qy	1498	ctgtctcttcacgt	1557
Db	2371	CTGTCTCTTCATCGCGCATCTGTCTTCAACGCTCTTCCTCCCGCAAGCCGCGAAGCGTGGCCG	2430
Qy	1558	aaccggtggggcggaattcgcgcgaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1617
Db	2431	AACCTACTGGAAACGACATCGCGACACGCTGTGAATGAGACCTCGCCCTCGCGCGCGCGAG	2490
Qy	1618	cacacgtctcgaagcgt	1674
Db	2491	CATACCTTCGAGACCTCTGCCAAGCGGAGACTGGGATGCGGCACACGCGCATTTGA	2547
RESULT	3		
PDCCOX2			
LOCUS			
DEFINITION		PDCCOX2 3211 bp DNA	BCT 07-OCT-1996
ACCESSION		P.denitrificans COI gene.	
VERSION		X05829	
KEYWORDS		X05829.1 GI:994806	
SOURCE		COI gene: cytochrome c oxidase; cytochrome oxidase a3;	
ORGANISM		unidentified reading frame.	
		Paracoccus denitrificans.	
		Paracoccus denitrificans	
		Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;	
		Paracoccus.	
REFERENCE			
AUTHORS		1 (bases 1 to 3211)	
TITLE		Raitio M., Jalli T. and Saraste M.	
JOURNAL		Isolation and of the genes for cytochrome c oxidase in Paracoccus	
AUTHORS		denitrificans	
TITLE		EMBO J. 6, 2825-2833 (1987)	
REMARK		2 (bases 1 to 3211)	
AUTHORS		Saraste M.	
TITLE		Submitted	
REMARK		Submitted (04-JAN-1991) Saraste M., EMBL, Meyerhofstrasse 1,	
AUTHORS		Heidelberg, Germany	
TITLE		revised by [3] MAT	
JOURNAL		3 (bases 1 to 3211)	
REMARK		Saraste M.	
AUTHORS		Direct Submission	
TITLE		Submitted (20-SEP-1995) Saraste M., EMBL, Meyerhofstrasse 1,	
COMMENT		Heidelberg, Germany	
		On Sep 21, 1995 this sequence version replaced gi:45474.	

FEATURES	source	location/Qualifiers
		1..3211
		/organism="Paracoccus denitrificans"
		/db_xref="taxon:266"
		/map="cox2 locus"
RBS		218..222
CDS		229..564
		/note="ORF4"
		/codon_start=1
		/transl_table=1
		/protein_id="CAA9273.1"
		/db_xref="GI:43475"
		/translation="MADISMPKASAAAGETSRAGASIDDAFLASKVVFADGALDKRPT KOLIAVAVAHYTQCPMCIECHVKGARREGATNEIMEIAIWAAPRAAGAAVAHANAKAL AVIDEIDGA"
		229..561
mat_peptide		/product="ORF4"
		628..640
repeat_unit		/note="inverted repeat A"
		686..689
TATA_signal		713..725
repeat_unit		/note="inverted repeat A"
		765..769
RBS		777..2444
gene		/gene="COI"
		777..2441
CDS		/gene="COI"
		/EC_number="1.9.3.1"
		/codon_start=1
		/transl_table=1
		/product="cytochrome-c oxidase"
		/protein_id="CAA9274.1"
		/db_xref="GI:995678"
		/db_xref="SWISS-PROT:P08305"
		/translation="MSA0ISDSIEKRGFFTRMFMTSTNKHDKGVLYLFTAGLAILSV TLTYVMRELQHPGVQYMCLESGMLVDAAECTPRNHLMMVVTYHQLIMLPFVIVL ALFPGFSGYEMPRHLIGADPMAPRLNNLSLYCYGSALASLSPGSDPQAVGYA WVLPFSTPEAGYAMDALITFAVVSQATSLGAINITITFLNMBARQMTFKYRLP WAVITRAMILSLPYLAGSTITMLMBRNCTQFFPDAGGSDPYLXOHLIMFGRNRE YMLLPGRITLIVSTFARKPIFGYLPVLAAMALAIFGLTVANHNHTKMSLDO TYFOMATNTIAVPGTIVFSWIAVMWGSLEFKTPIMLALFLFTVGAVTVVAIDQGS LDRYAHDTYIVAHFHYVMSLGALFAIFAGTYVIIGKMSGRQYDEMAOHLFNMKFI SNLIFPFOHFGRGQMPRRYIDYVEFSEYNNNISIGAYISFASFLFICIVFTLFE GKPVNPVYMNVEHADLTLEWTLPSPPREHTFETLPKPEDMDRAQNR"
		777..2444
mat_peptide		/gene="COI"
		/EC_number="1.9.3.1"
		/product="cytochrome-c oxidase"
		2744..2755
repeat_unit		/note="imperfect inverted repeat B"
		2744..2788
terminator		2764..2775
repeat_unit		/note="imperfect inverted repeat B"
		2764..2775
BASE COUNT	578 a 1008 c 998 g 627 t	
ORIGIN		
Query Match	59.2%	Score 991; DB 1; Length 3211;
Best Local Similarity	75.6%	Prod. No. 9.6e-198;
Matches 1244; Conservative	0; Mismatches 395; Indels 6; Gaps	1.1
QY	11 ccgcattcacagcgcatgacacacatlgaaagaagaagccttcttcacgcgcgtgttcaatgt 70	
DB	781 CAGCCCAAGATCAGCGACACCATCGAAGACAGCGGGCGTTCTTACCCCGCGTTCATGT 840	
QY	71 ccgacacacacaaagacatcggtctgctataacctgtaagcgcgctgtgtgtttca 130	
DB	841 CCACGACACACACAGATATCCGCTGTCCTTACTCTTCACGCGGGGCTTCGCGCGGCTCA 900	
QY	131 tttcgtctgttaccgcctcatatagcgcttgagcgttgatgataccgagtggttca 190	
DB	901 TCTCGGTCACTACCGCTCTACATCGCATGGAATTCAGCATCCGGGTGTGCATATA 960	

QY 191 tgcgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 250
 Db 961 tctgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 1020
 QY 251 accgtggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 310
 Db 1021 tctgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 1080
 QY 311 cggagcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 370
 Db 1081 tctgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 1140
 QY 371 tggccttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 430
 Db 1141 tggccttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1200
 QY 431 gctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 490
 Db 1201 tctgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 1260
 QY 491 tctgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 550
 Db 1261 tctgcttgaagcgacgctctgacgagatgctcgcagacatgtaacgagacgac 1320
 QY 551 cggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 610
 Db 1321 cggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1380
 QY 611 tgaacatgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 670
 Db 1381 tgaacatgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1440
 QY 671 ttaacatgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 730
 Db 1441 ttaacatgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1500
 QY 731 tgccttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 790
 Db 1501 tgccttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1560
 QY 791 tctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 850
 Db 1561 tctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1620
 QY 851 cggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 910
 Db 1621 cggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1680
 QY 911 accgtggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 970
 Db 1681 accgtggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 1740
 QY 971 accgtggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 1030
 Db 1741 accgtggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 1800
 QY 1031 tggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1090
 Db 1801 tggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1860
 QY 1091 gctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1150
 Db 1861 gctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1914
 QY 1151 tggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1210
 Db 1915 tggcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1974
 QY 1211 cctactatgctgagcagatgctgacgagatgctcgcagacatgtaacgagacgac 1270
 Db 1975 cctactatgctgagcagatgctgacgagatgctcgcagacatgtaacgagacgac 2034

QY 1271 tgcgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1330
 Db 2035 tgcgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 2094
 QY 1331 caagctggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 1390
 Db 2095 caagctggaagcgaatgctgacgagatgctcgcagacatgtaacgagacgac 2154
 QY 1391 tctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1450
 Db 2155 tctgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 2214
 QY 1451 ggaacatgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1510
 Db 2215 ggaacatgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 2274
 QY 1511 tgaacttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1570
 Db 2275 tgaacttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 2334
 QY 1571 aatgcgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1630
 Db 2335 aatgcgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 2394
 QY 1631 cgtgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 1690
 Db 2395 cgtgcttgcgagatgctgacgagatgctcgcagacatgtaacgagacgac 2450

RESULT 4
 LOCUS RSCOX1
 DEFINITION R.sphaeroides gene for cytochrome oxidase subunit 1.
 VERSION X62645.1 GI:46417
 KEYWORDS cytochrome oxidase subunit 1.
 SOURCE Rhodobacter sphaeroides.
 ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.
 REFERENCE 1 (bases 1 to 2166)
 AUTHORS Shapleigh, J.P.
 TITLE Direct Submission
 JOURNAL Submitted (17-OCT-1991) J.P. Shapleigh, University of Illinois, Dept. of Biochemistry, 505 S. Mathews, Urbana IL 61801, USA
 REFERENCE 2 (bases 1 to 2166)
 AUTHORS Shapleigh, J.P. and Gemmis, R.B.
 TITLE Cloning, sequencing and deletion from the chromosome of the gene encoding subunit I of the a33-E type cytochrome c oxidase of Rhodobacter sphaeroides
 JOURNAL Mol. Microbiol. 6 (5), 635-642 (1992)
 MEDLINE 92204019
 FEATURES
 source location/Qualifiers
 1. 2166 location="Rhodobacter sphaeroides"
 /organism="Rhodobacter sphaeroides"
 /strain="GA"
 /db_xref="taxon:1063"
 207. 1904 /transl_table=1
 /product="cytochrome oxidase subunit 1"
 /protein_id="CAA44514.1"
 /db_xref="GI:46418"
 /db_xref="SWISS-PROT:P33517"
 /translation="MADAAIRGHEHRRFETRMWSTNRHNDIGVLYLFTGGLVGLIS
 VAFIVYRMLAIPGVDFWCAHLESGVLFKGFQSLWPSAVENCTPNGHLMVMTGTH
 GILMFVYIPALFGFNGYFMPHLIGADMAEPKNNLSVLYVAGSLAVASLAP
 GGNGQSGSGIGWLYPLSTSSGVDIAIFAVHLSGASLLGA1NM1TTPFLNRRAP
 GMTMKVPLFAMSIFVTAMLLIALPVLGATLMLTDBSGTFFQPSGGDDPYLYO
 HLMRGHEVYIVIPARGIVSHVATPAKPDITGILPMYAMAYICVGLGVVNAH
 MTAGSLTQOSTFMATKIVAVPISIKIFSLIATIMWGSIELKPMIALSGELPLTF
 VGGVTVISQASVDRYVDHYVVAHVFVSLDAVPSIFAGSTNSGICKMSGRQYPE
 MGKLFHFMFVGANLTFPPHFLGRQMPRYIDYPEAFATWNVSSILGALSPASF

PUBMED 11481430
 REFERENCE 2 (bases 1 to 286550)
 AUTHORS Gouzy J.
 JOURNAL Direct Submission
 COMMENT Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO EU Consortium
 MELILO EU Consortium:
 Laboratoire de Biologie Moléculaire des Relations
 Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
 France, Laboratoire de Génétique et Développement UMR6061-CNRS,
 Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,
 France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
 Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
 D-33615 Bielefeld, Germany, Unite de Biochimie physiologique 20,
 Université catholique de Louvain, Place Croix du Sud 2, Bte 20,
 B-1348 Louvain-la-Neuve, Belgium, Unité de Microbiologie, Faculté
 des Sciences Agronomiques de Gembloux, Avenue Mareschal Juin 6,
 B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
 http://sequence.toulouse.inra.fr/meliloti.html.
 FEATURES
 source
 1..286550
 /organism="Sinorhizobium meliloti"
 /strain="1021"
 /db_xref="taxon:382"
 70..813
 /gene="gstII OR SMC00916"
 70..813
 /BC-number="2.5.1.18"
 /function="miscellaneous"
 /note="Product confidence : putative
 Gene name confidence : putative
 Predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
 /protein_id="CAC45361.1"
 /db_xref="GI:15073720"
 /translation="MADLSFPTTARKPAKNPDIOLYSLPENGKVKVSTALEELGIA
 YEHRISFDNEKSPESLNPNGRIPIALIDNGPKPIGLFESGAILIYAETGT
 KLIPADARREYETLCWFMFGMGIGPFGHFFKADKANNPNYPVORYSDEAKR
 LILVEARLEGRWLKMGNTPTSPSTPIPGWGRSSSVVAPRCSTLRASRMSTGMI
 AAMHALPBRARMSRARSRARSLPAGD"
 588..779
 /gene="SMC00915"
 588..779
 /gene="SMC00915"
 /function="miscellaneous: hypothetical/partial homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 Predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="HYPOTHETICAL PROTEIN"
 /protein_id="CAC45362.1"
 /db_xref="GI:15073721"
 /translation="MADGGEYTIADIATYVWEGARKFYGAELVDYKSPNMDV
 DRGLRPAAGKGMETPRKE"
 867..2084
 /gene="SMC00914"
 867..2084
 /gene="SMC00914"
 /function="small molecule metabolism"
 /note="Product confidence : putative
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="HYPOTHETICAL PROTEIN"
 /protein_id="CAC45363.1"
 /db_xref="GI:15073722"
 /translation="MSGRLVVGCGGAFAFVAKLRALKDKMRPTTVAAEASLPYORP
 PLSKYLRLEMTLDRLYRPEAWYAEHEIDRLSTTVYVDRLAQVALSDSMLTYE
 TLAFATGAPRRRLPAVGGDLAVGRDPRADRLAEEMGRRVLVGGGYIGLEA
 AAVAFSGEAGTYEMADRIILORVSAATSAIVREIHRSHGVDIRRIGRLIGDYG
 RYTAELSGSVYIPDVIYVIGVANAADLAHEAGIETRANGIVDSHSGTSPTTYAM
 GGCACVPMGMRIRLESVYMANADQAEAVAVIAGGTDPIIDPKPWRMSQIDYKDLIAG
 FGIHDETLVROGQROGSVSWVFROGKLIVADAINDAKAVYTKLLKAGATPDRLL
 LADPYVDLALIV"
 complement(2114..2572)
 /note="Sm-2 OR SMC04876
 REPEAT SM-2
 predicted by Homology"
 /evidence=not_experimental
 complement(2585..4222)
 /gene="groEL1 OR SMC00913"
 complement(2585..4222)
 /gene="groEL1 OR SMC00913"
 /function="cell processes: chaperoning"
 /note="Predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="60 KD CHAPERONIN A PROTEIN"
 /protein_id="CAC45364.1"
 /db_xref="GI:15073723"
 /translation="MAAKEVRFGRSAREKMLRGVDILADAVKVTLGPKGRNVIDKSR
 GAPRITKIDGVSAKEIELEDEKFNMGAGQREVAKKINSDAEVAGGTISANGROIGL
 GAKAAVAGNPMDLKRGIDLAEEVYKDLAKRKINSDAEVAGGTISANGROIGL
 DIAFMOKVNGSVITVEBAKTAETLEVEVMOPDRGLTSYPTNPERKVADEDA
 FLIHRKISNQAMHPIYEAIVGTGKPLITIAEVEBEALATLVNKLKGLKIAAV
 KAPGFDERRKAMLEIDITLGTGTVISDEGKILESVDLMDLGRARVITKEENTIVD
 GAGQKSDIEGRAVQIKQILETTSDYDEKLOERLAKLAGGAVYRVGAGVEYKEK
 KRIDIDALNATRAAVQEGIVPGGVALLRSSVYITKGEENDODAGVNIYRRALQSPA
 RQIVNAGSEASIVVGIILEKNTDDEFGVNAQGEYGDVMIANGIIDPVVAVRTALQDAA
 SVASILITTEAMIAELPKKAPAMGGMGCMGMDMM"
 complement(4298..4594)
 /gene="groES1 OR SMC00912"
 complement(4298..4594)
 /gene="groES1 OR SMC00912"
 /function="cell processes: chaperoning"
 /note="Predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="10 KD CHAPERONIN A PROTEIN"
 /protein_id="CAC45365.1"
 /db_xref="GI:15073724"
 /translation="MASTNFRPLHDVVVVRVSESEKTKGIIIPDTAKERPQGEIV
 AVSGSARDSESGVPLDVKAGDRILFGKWSGTEVKINEDLLIMKEMDIMGVIG"
 4863..5690
 /gene="SMC00911"
 4863..5690
 /gene="SMC00911"
 /function="miscellaneous: hypothetical/global homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental

predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="PUTATIVE OXIDOREDUCTASE PROTEIN"
 /protein_id="CAC45363.1"
 /db_xref="GI:15073722"
 /translation="MSGRLVVGCGGAFAFVAKLRALKDKMRPTTVAAEASLPYORP
 PLSKYLRLEMTLDRLYRPEAWYAEHEIDRLSTTVYVDRLAQVALSDSMLTYE
 TLAFATGAPRRRLPAVGGDLAVGRDPRADRLAEEMGRRVLVGGGYIGLEA
 AAVAFSGEAGTYEMADRIILORVSAATSAIVREIHRSHGVDIRRIGRLIGDYG
 RYTAELSGSVYIPDVIYVIGVANAADLAHEAGIETRANGIVDSHSGTSPTTYAM
 GGCACVPMGMRIRLESVYMANADQAEAVAVIAGGTDPIIDPKPWRMSQIDYKDLIAG
 FGIHDETLVROGQROGSVSWVFROGKLIVADAINDAKAVYTKLLKAGATPDRLL
 LADPYVDLALIV"
 complement(2114..2572)
 /note="Sm-2 OR SMC04876
 REPEAT SM-2
 predicted by Homology"
 /evidence=not_experimental
 complement(2585..4222)
 /gene="groEL1 OR SMC00913"
 complement(2585..4222)
 /gene="groEL1 OR SMC00913"
 /function="cell processes: chaperoning"
 /note="Predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="60 KD CHAPERONIN A PROTEIN"
 /protein_id="CAC45364.1"
 /db_xref="GI:15073723"
 /translation="MAAKEVRFGRSAREKMLRGVDILADAVKVTLGPKGRNVIDKSR
 GAPRITKIDGVSAKEIELEDEKFNMGAGQREVAKKINSDAEVAGGTISANGROIGL
 GAKAAVAGNPMDLKRGIDLAEEVYKDLAKRKINSDAEVAGGTISANGROIGL
 DIAFMOKVNGSVITVEBAKTAETLEVEVMOPDRGLTSYPTNPERKVADEDA
 FLIHRKISNQAMHPIYEAIVGTGKPLITIAEVEBEALATLVNKLKGLKIAAV
 KAPGFDERRKAMLEIDITLGTGTVISDEGKILESVDLMDLGRARVITKEENTIVD
 GAGQKSDIEGRAVQIKQILETTSDYDEKLOERLAKLAGGAVYRVGAGVEYKEK
 KRIDIDALNATRAAVQEGIVPGGVALLRSSVYITKGEENDODAGVNIYRRALQSPA
 RQIVNAGSEASIVVGIILEKNTDDEFGVNAQGEYGDVMIANGIIDPVVAVRTALQDAA
 SVASILITTEAMIAELPKKAPAMGGMGCMGMDMM"
 complement(4298..4594)
 /gene="groES1 OR SMC00912"
 complement(4298..4594)
 /gene="groES1 OR SMC00912"
 /function="cell processes: chaperoning"
 /note="Predicted by Codon usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="10 KD CHAPERONIN A PROTEIN"
 /protein_id="CAC45365.1"
 /db_xref="GI:15073724"
 /translation="MASTNFRPLHDVVVVRVSESEKTKGIIIPDTAKERPQGEIV
 AVSGSARDSESGVPLDVKAGDRILFGKWSGTEVKINEDLLIMKEMDIMGVIG"
 4863..5690
 /gene="SMC00911"
 4863..5690
 /gene="SMC00911"
 /function="miscellaneous: hypothetical/global homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 predicted by Codon usage
 predicted by Framed"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental


```

      1 1111 1111111111111111 1111 1111111111111111 1111 1111
Db 128002 AGGCCCAACAGCGTGAATGACGACCTCTCTCCGCCGCCGTTCCACCACTGGACGA 128061
QY 1632 gctgcaccaagcgctcgactgggacaagca 1661
      1111111111111111111111111111111111111111111111111111111
Db 128062 GCTCCCGCGGATCAAGTACGCGGTAAACGAA 128091

RESULT 6
LOCUS AE006001/c 11188 bp DNA BCT 28-MAR-2001
DEFINITION Caulobacter crescentus section 327 of 359 of the complete genome.
ACCESSION AE005673
VERSION AE006001.1 GI:13425117
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 11188)
Niernan, W.C., Paldylyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J.,
Heldelberg, D.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I.,
Potocka, A., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
Ely, B., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
Helft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
Emolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and
Fraser, C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
2 (bases 1 to 11188)
Niernan, W.C., Paldylyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,
Heldelberg, D.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I.,
Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
Laub, M.T., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
Helft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
Emolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
Fraser, C.M.
Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
location/Qualifiers
1. 11188
/organism="Caulobacter crescentus"
/db_xref="taxon:69394"
complement(94..1752)
/gene="CC3406"
complement(94..1752)
/gene="CC3406"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="cytochrome c oxidase, subunit I"
/protein_id="AAK25368.1"
/db_xref="GI:13425118"
/translation="MASADTDHAAHDDHKKPEFARWPFSTNHHKDCITLYTIFATMA
GIYGGALSLIRWELMAGPQOVSDGMLAOMCIFKSGICYNNAVTAHALIMFFVVM
PAIIGGCGNFVIMIGAPDPAWIRKNNISFWILLIAWILLVSMFVDDGPGGFEFGG
WTLYPLSTGTHGPMADLAISLHLAGASSIIAGINFTTIFNMRAPGTLHMPLEF
VWSVLITAFLLLSLPLVAGATIMLTDRNFGTHFPDPAAGDPVAFMHOILFWFGHDE
VYLLIFGFGIISHIVSTESKRPVGYLAAYMAVIAIGFVYVMAHHVTVGMSINL
RAYEVAVMIIAVPTGKIRSWIATMGSGLSFRTPMNAIGFIFLFTVGGVGVYS
NAGIDVTLHDYVVAHHVHVLVSLGAVFAFAFYVWFEMKGVKNEPELGAHHFVIM
FIGVNLIFRPHFLGLGGMRRKRVDPDAITLNNVSTVGMITTVGVGVFMVLIEA
AIRRKAANPWESEGAFTLEWTLSPPPHQFSEPEVIGDDHH"
gene
CDS
complement(1779..2708)
/gene="CC3407"
complement(1779..2708)
/gene="CC3407"
/note="identified by match to protein family HMM"
/codon_start=1

```

```

/transl_table=11
/product="cytochrome c oxidase, subunit II"
/protein_id="AAK25369.1"
/db_xref="GI:13425119"
/translation="MKAQSGIRIRIMTGASALATMMFAGHALAEDLMGOTPGAIDL
OPASELKHAIWFHDMILPTITATLVLGLLIMIVRYKKKANPAKAFNSHTTI
EYMTVYVPLIIMVIAIFSERLFHNNPKRDLTVKATGYMYNGYEPPOKSEII
SLPLIKQAEAKKVPITLADANAVYVANKVVRVOVITADVIHAFILFGLKIDAI
GRINTEWFKADRTGVFYGGSELGVDHAFMPETIKVYVQAEFDMAVAKTAPAPAA
AAEAPAAAGCATPASAPAAVASVPAAPAPAA"
3083..3409
/gene="CC3408"
3083..3409
/gene="CC3408"
/note="identified by glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK25370.1"
/db_xref="GI:13425120"
/translation="MPETIEIQLKRGVIALCYLALLSHADSYAEIASRLTRDIDMG
GTYPLMRMQSDGIVETLVBSQSGPPKYYRLTAAGRESIAQKAEWTAFAVAVN
IVGAAA"
3406..3969
/gene="CC3409"
3406..3969
/gene="CC3409"
/note="identified by glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25371.1"
/db_xref="GI:13425121"
/translation="MTRAEITRLRKGSLGSLPATTIADIADHEAFSDAQAGRTA
EVAALGDDRDRLARELRABAGLKEWEEKTPSNAAAIITALICAIIDVILPLIMG
VYATLFGFLKVAVTTFVFGGFIFAAGPLAAPGPGITFALGIGIMASATFTSLTL
VIVGVFNAAVWVAIRHLRFLKALPQA"
3966..4802
/gene="CC3410"
3966..4802
/gene="CC3410"
/note="identified by glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK25372.1"
/db_xref="GI:13425122"
/translation="MSREGMILIRNTLIVAGASPALAVGCFAGVAAIAGPELMTKGM
TIPGDSASVTTDDNGKVIITVRHVAGAGAPWDRTLPTGKDTLIDLPDVFV
VOGAERKIVSGPKSYVDRLKVEGGRISLIDNGAIDMTLGPGLGVRSDBRMKI
TVAPDVVPEFVEYRSGGDLDIRYBOALINDVSSGDYEAEGKVFYTVVENSQSYAD
LQDLKAREALITEVSSGSGAVATATKAKIDISGSDVELRTPEKSVTSETITSGEVR
ET"
4964..6436
/gene="CC3411"
4964..6436
/gene="CC3411"
/note="identified by match to PFAM protein family HMM"
PF01523"
/codon_start=1
/transl_table=11
/product="tld protein"
/protein_id="AAK25373.1"
/db_xref="GI:13425123"
/translation="WTSAFGSLTCKTMNAPIPASSLITLDAAGVDDPRARAIIIGBALA
GADDELPLERSESEAFVDDDRKLSASVDSGEGLVNAGVAGETVAAHAEISEAI
IRASDSASLAKGHAGVSAEGRATNEKLYGDNVSAPESDKVALLOEIDARAR
DPRVOVNASMAGSERVVEILRAOGRILRDVAPLVRAVVOYTVKEDGRRSGSAGG
RAGFAWISPKMDQOVDEALMALVNLDAVACPGEMDVYLGGMNVYLIAHVAHGS
LEGDPNRKGISFSGRIGERYAABGVYTFYDGSIPGRGSLTVDDGETPTERTTILIED
GLIVGIMDRKMSARLMGKATGNGRSYAHMPMRMTNTGLAGEHHPQEMIASLKR
GLYCANFGGQVDITNGKVFVQCPEAYLVEBGKITAPVKATLIGDPSALTIRVTMIG
NDPDPGIGVCGKSGGCVPGVGPISIKITGLITVGTAV"

```

```

(gene) complement(6601..7098)
CDS /gene="CC3412"
      complement(6601..7098)
      /gene="CC3412"
      /note="identified by match to protein family HMM"
      /codon_start=1
      /transl_table=11
      /product="acetyltransferase, GNAI family"
      /protein_id="AAK25374.1"
      /db_xref="GI:13425124"
      /translation="MSRLSLPLADLADNAPGALLALNNAHVELSLMEPERLAHLV
      AQAFLARVRVADALLTFEDHDADSPNFMFERRPYFVYVDRVVADSAKRGSLA
      RALYDLDFDAAKAAGHARIVCEVNSDPNPVSDAFHALLGEFVPGTAEIHGKKTVY
      LERRL"
(gene) complement(7386..9266)
CDS /gene="CC3413"
      complement(7386..9266)
      /gene="CC3413"
      /note="identified by match to protein family HMM"
      /codon_start=1
      /transl_table=11
      /product="peptide transporter"
      /protein_id="AAK25375.1"
      /db_xref="GI:13425125"
      /translation="MNIIVAGILVLTGIPVLQLKHPRLIICFLAEMERES
      YYKRAIILFYLTQHFDPKASGHSYALVLPILIGFLADKMLGTRKAVAF
      AVLAAGHLAMIEGKPAVOITLYOGATVEQAEGRGNREKLVDDKALEVATK
      GPEIKGLPANAPLPAVLPAKADYKLDVAGRPDLINIFMSLSIILIGVLEKPIST
      IYGLYKPGDPRDRPGLTYXGILGSPMASIIGLIGLVNVMAGGALGVGMLAG
      FIVFLGKRWLDGKAEPPEPKYLKRVAGPIRELAIYGLSLVGLVGFPIYORALY
      GVALLNGILASLGIYGFEMFTKCAEERELAVFLIFGAVETWLEEQAGSSLSLF
      AATNOLDLVRPLRLFSGAVTLAPDOLAAGIDAFTFVNDTSFNAPQOALNAG
      ILIFAPVAWVTFGLARGKDPNPVIRKGLALIQYAGFLLDNGSTPADGAFKPLPI
      FLVLMYHTTGEICMSPVLSOMTKLSPASVSPMAWVFLAVAIQWVGSKLAGLA
      ATEYVGGVLDGGAALASIKYFNITGLVAVALIGFLLSPVLKKNHGDVNGAPK
      KA"
(gene) complement(9726..11087)
CDS /gene="CC3414"
      complement(9726..11087)
      /gene="CC3414"
      /note="similar to GP:3688156; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="glutamate-cysteine ligase"
      /protein_id="AAK25376.1"
      /db_xref="GI:13425126"
      /translation="MADTRASVOEROLITLDTLTAFAKSKPEARVGAHEHKKGYL
      GSHEPVYEGDKGVHALLGLDORFGMTPEMSPITIGLERNCANVSLRPGQFELSGA
      PLSTMEICEETGQHLDEVKTVADDELIGFTGLGSPKWTQVPIPRGRVIMRNY
      MPKVNGLIGDMMRLRTVOANDLPSSEADMAKFPMSIALDPIATLFRANSFTDEGK
      NSELSARAVWTFDTPNRTGLDFVEEDGFDEERYARVALDPIFVIRGDKYIDVAG
      RSFRDPIEKLPELPELATKDFADHTTAPREARLTYLMEKRCADGPMRSKALP
      ALMTGVFYNDALAAAMDLCRMTLEDREGLARDPKLGIAQVAGRAAOYAKRFA
      VAKGLKNAQLNGGFLEDTITLUGLELOLADSGITPAEKLSLYNGANGDIERVFTD
      VAK"
BASE COUNT 2025 a 3680 c 3749 g 1734 t
ORIGIN

```

Query Match 44.8%; Score 750.2; DB 1; Length 11188;
 Best Local Similarity 67.5%; Pred. No. 2.2e-147;
 Matches 1122; Conservative 0; Mismatches 518; Indels 21; Gaps 4;

```

Db 1620 GGCATCGCGGTGGCGGCTTTCCGGCCCTCATCGCGTGGAGCGATGAGCGCGGCGATT 1561
Qy 184 cagtaacatgctctgaagcgacgctctgatcgatgctctgcgaatgaagcg 243
Db 1560 CAGGTGTTCTCCGACACCGG-----CATGCGCGCCGATGGCATCTTCAAG 1513
Qy 244 aacggacacccgctggaacgcatatgatactacatgatacttcatgatacttctgtg 303
Db 1512 GGCAGACGCGGCTATMAACGCGCTGTGACGCGCCAGCCCGCATGATCATCTTCTCATG 1453
Qy 304 ggtatcccgcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 363
Db 1452 GTCATGCTTCGATGATGATGCGGCGGCTTGGGCAACGATGTTCTGTCATGATGCGGCG 1393
Qy 364 ccggaatgagctcccgcccgatgaacaaactgtctgtctgtctgtctgtctgtctgtctgt 423
Db 1392 CCGGACATGCGCTTCCCGGATGACACATCTCTGTTGCGTGTGATGCGCGCTCG 1333
Qy 424 gcatgagcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 483
Db 1332 ATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1276
Qy 484 ggttggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
Db 1275 GGTGGAAGATCTATCCGCGCTGTGACACATGCTGTCAAGGGTCCGCGCATGAGACTG 1216
Qy 541 ggaattccgagctgaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 600
Db 1215 GCGATCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1156
Qy 601 aagacattctgaacatgacgagcgcccgacgaacgaacgaacgaacgaacgaacgaacga 660
Db 1155 ACCAAGATCTTCAATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1096
Qy 661 tgttcgactcttaccagcgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
Db 1095 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1036
Qy 721 atcaccaatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
Db 1035 ATCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 976
Qy 781 ggtgacccgactctgttaccacacacacacacacacacacacacacacacacacacacac 840
Db 975 GCGCGACCCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 916
Qy 841 atcattcgcgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
Db 915 CTGATCCTGCGCGGCTTGGGCAATCATGACCATCATGCTGCTGCTGCTGCTGCTGCTGCTG 856
Qy 901 gcttcgacttaccbctgcgactgtctatgaatgggtgtgtgtgtgtgtgtgtgtgtgtgt 960
Db 855 GCTTCTGCGCTACCTGCGGATGCGGCTGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 796
Qy 961 gctcgtgacacacacacacacacacacacacacacacacacacacacacacacacacac 1020
Db 795 GTGTGGGCGCACACATGATGACATGACATGACATGACATGACATGACATGACATGACATG 736
Qy 1021 ctggccacacatgatactgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
Db 735 GCGCGCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 676
Qy 1081 atgttgagcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
Db 675 ATGTGGGGGGTTCCTCAGGTTCAAGACGCGCATGCTGTGGGGGATGCGGCTTCATCTTC 616
Qy 1141 ctgttcacccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
Db 615 CTTTTCACGCTGTGGGCGGTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556
Qy 1201 tatcagacacacatgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260

```



```

Db 555 CTGCACGACACCTACTACGTGGTGGCGGCACTTCACTACAGTGTCTGTGCTGGCGCGGTG 496
Oy 1261 ttgtgacatttcgacgagatcttcaactttcaatgcgaagttctcgagccgcttcccg 1320
Db 495 TTGCGCATCTTGGCGGGCTTCTACTACTGTTCGACAAAGATGTGGCGCGTGAAGTACAAAC 436
Oy 1321 gaatggatgcgaagaactgcgtacattcgtacattcattcattcgtgcgaagtcacgtcttc 1380
Db 435 GAGTTCCTCGGTGGCGGCCACACTTGTGATCATGTTCATGGGTGTGACACTGATCTTCTTC 376
Oy 1381 ccgcagcacttcctgagcagttcaaggatgacgcgcgcgttacatcagactaccccaagcc 1440
Db 375 CCGCAGCACTTCTCTGGCGCTGCAAGGCAAGCCGCGCTATGTGCACTATCCGAGCGCC 316
Oy 1441 ttcgcgcgttgacaagaatctcgtccatggtgctgccttcctgacctgcccctgcttcg 1500
Db 315 TTACACCTGTGAACTACTGTCGACGGTCGCTCATGATATACCGTCTGTTGTTGCGGC 256
Oy 1501 ttcttcacgtatcttctgtctatacgtctgttgcgtgcgcgcgaagccgctccgaac 1560
Db 255 GTGTCATGCTGTGCTGATCGAGGGCGCGAT---CCGTCGCGCGCAAGGCCGAGCGCAAC 199
Oy 1561 ccgtggcgcgaatttcgcgcgatacgtgaaatgagcgtgcacacgccttcgcgcgcac 1620
Db 198 CCGTGGGCGGAAGGCGCCACCAAGCTGGAATGACTCTCTCTCGCGCGCGCGCTTCCAC 139
Oy 1621 acgttcgaacgcgtgcacacgcgcgtgcgactgagacaaga 1661
Db 138 CAGTTCACGCAACCGCGGTGATCAAGGGCGACGACCA 98

RESULT 7
AE008010 10835 bp DNA BCT 14-AUG-2001
LOCUS Agrobacterium tumefaciens strain C58 circular chromosome, section
DEFINITION 68 of 254 of the complete sequence.
ACCESSION AE008010 AE007869
VERSION AE008010.1 GI:15155740
KEYWORDS
SOURCE
ORGANISM Agrobacterium tumefaciens.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE
1 (bases 1 to 10835)
Hinkle,G., Slater,S.C. and Goodner,B.
Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
Unpublished
2 (bases 1 to 10835)
Hinkle,G., Slater,S.C. and Goodner,B.
Direct Submission
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
source
1..10835
Location/Qualifiers
/organism="Agrobacterium tumefaciens"
/strain="C58"
/db_xref="taxon:358"
/complement(118..642)
/gene="AGR_C_1396"
/complement(118..642)
/gene="AGR_C_1396"
/note="(AJ242592) hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1396p"
/protein_id="AK86575.1"
/db_xref="GI:15155741"
/translation="MRLSPRLRTFLAAGLAIIPAAALAAQOPPGPKSTHGAWSY
ICDKPAGASEOCALMONVITADDPKGLSVYITKTDPRKSRITRIILAPGLKDGK
ELIYDNNNIGATYTRCFSEGCYIEVDIDELLVLRAGNAVPALRESVDQDRVGI
IETSGFAEGYDALP"

gene
988..1938
/gene="AGR_C_1397"
988..1938
/gene="AGR_C_1397"
/note="(AJ242592) cytochrome c oxidase subunit II"
/codon_start=1
/transl_table=11
/product="AGR_C_1397p"
/protein_id="AK86576.1"
/db_xref="GI:15155742"
/translation="MDSASVIDLDOTREGCVTNRIRYAAALAMTCLLTAVGAHADOP
VHMOMGMOEATPIHMEIRMFQYTLMTFVPTLVALLIIVAKFPAKRPVSKT
SHNTAIEVYWTAPVLILFLAFPSFNLLAQIDPENPDLTKATATQMLMSYKTA
AEGAPLSPDSYLLKDDPRAAGKEDKARYPRLLAVDNEMVYVCKTYRLTYAAPTD
VIHAFAMPAFGKIDAVGRLNETWFKPEKECLTYGCSSELCGKHAFPIAIRVSE
DOYNTWHAASADLNGANRALMASVDGARIVDAANETN"
2549..3664
/gene="AGR_C_1399"
2549..3664
/gene="AGR_C_1399"
/note="CYTOCHROME C OXIDASE POLYPEPTIDE I (CYTOCHROME AAS
SUBUNIT 1)"
/codon_start=1
/transl_table=11
/product="AGR_C_1399p"
/protein_id="AK86577.1"
/db_xref="GI:15155743"
/translation="MDLAIFSLHVGASSITGAINFITTLNMRAPGWTLHKMPLFPAW
SVLVAFLLSLPLVLAGITMLLDNRNGTAFESPEGGDPILOHLEMFEGHEVY
ILILPFGILISHIVSTPSFKPYEGLNAAVAVAGVFIWAHMYTVGLSLDAOR
YFEPATNVIAPVTGKIKIFSMITMGGSLTFSTPMWAIIGFLFTVGGVTVOLANA
GLDRSLHDPTVYVAHNVHVLSTGAVFAIFAGVYVMPKPTGYNTGNTVEFLGKLPWFPI
GVNLFFPQHFLGLAGMERRYIDYDPAIAGNMVSYSGYSIAVAVGIFLFEVWFAFA
KKRIAGNMPWGGATLLEMLSSPPYHOMEDLPRI"
4054..5007
/gene="AGR_C_1402"
4054..5007
/gene="AGR_C_1402"
/note="(X89566) heme O synthase"
/codon_start=1
/transl_table=11
/product="AGR_C_1402p"
/protein_id="AK86578.1"
/db_xref="GI:15155744"
/translation="MTVIDDRMGAESESLSEACARDYFELAKPRWSLVYFTAFAG
LVLAGETINPLIGLAIICIAVGAGSALNKNWYADIDAVMTTRAKRPISGRIPR
EALAPGLTSAFSAVILGLAVNMFSAAGLAFIPFVAVYVIMLRSTPONTVIGGA
GAFPPMLGMACVTGVSLSVILFLIIFLMTFPAHMAVALFEMRDRGAVGIPMPNVA
GERSTKNOMIYAVVITAAAVAPFPTGLASAGYGFPAVLASIFVYCSQEDVRMPEGD
EKMLPAKKMPAVSVILFALFISGLADHFAPAKAVISGL"
5167..5784
/gene="AGR_C_1406"
5167..5784
/gene="AGR_C_1406"
/note="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
/codon_start=1
/transl_table=11
/product="AGR_C_1406p"
/protein_id="AK86579.1"
/db_xref="GI:15155745"
/translation="MANMEAEATARGSORVNSRILVLCVFCMIGAMAYAAVPLYS
LECRVYNGTTOREYOYSDVILDKTIVTIDGANSNGLNMDRFRVDKMKVPKGETI
KVFPAKTRSPVATGTAVFNTVPEACAYANKVCRCFETETTLQDGETLEMPVVFV
DPDITARETYSVHTLITSYTFYPAKTEKPVASLPVKTESGSKL"
5846..6721
/gene="AGR_C_1407"
5846..6721
/gene="AGR_C_1407"
/note="(AJ242592) cytochrome c oxidase subunit III"
/codon_start=1
/transl_table=11
/product="AGR_C_1407p"
/protein_id="AK86580.1"

```

```

/db_xref="GI:15155746"
/translation="MADTHOKNHDYHIIDPSPWPLLASIGAFIMTGGVCVMRLTSG
SPKLEGLANLWPLFYGLVLYVMYAMWADTIKEANESGSHTRVSLHLYGIMETI
ASEVWFPAWEHAYFADASLREPEAIOASRLLEYTGQMPKCEIVIDPMLHLYNTVIL
LISGCVTAHHAHLHNDKGLISGLATLVALGVLFSTVOYEITHAFEDKRNSTYGA
TFPMATGHEHVEVTEVLLVCLFRAIAGETTPKHGFBAAMWYHVEDVWMLFLR
FAIYIMGGMGAPLH"
6941..7369
/gene="AGR_C_1409"
6941..7369
/gene="AGR_C_1409"
6941..7369
/gene="AGR_C_1409"
/translation="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1409p"
/protein_id="AAK86581.1"
/db_xref="GI:15155747"
/translation="MYSRFGCQREHREVLMSPESSNGNFAPVDPTKVLGKCCPCGNGR
LEDFGLTRPACSGLDYGFADAGDPAVEFMILVGLVGMALWMDORPAPVMAH
VMLMPEFYIVSLVLRKLGIMIALQYRNNASEGRDRE"
7338..8114
/gene="AGR_C_1411"
7338..8114
/gene="AGR_C_1411"
/translation="Surfeit 1"
/codon_start=1
/transl_table=11
/product="AGR_C_1411p"
/protein_id="AAK86582.1"
/db_xref="GI:15155748"
/translation="MQAKGLIVNDVSHMKAQRWWEAPLYLLALAILLGTMQY
KRLYKEMALMADIEERRNAPATLSIDIAIVSGCEIEYRKRLSGTDHREIRFEA
THOGGVYITPLFLADRLIFVNRGEVPEPMKAARPDGOVGEVYITGLAAPL
VAKPSLLPDNDIAKNIEFYKMDLAAMASAEVPERIKLFDAANAPNPGMPQGGV
TILDPNNHLOQYAITWYGLAALVIVAGFAFRNKAQGE"
8173..9219
/gene="AGR_C_1414"
8173..9219
/gene="AGR_C_1414"
/translation="Lyb"
/codon_start=1
/transl_table=11
/product="AGR_C_1414p"
/protein_id="AAK86583.1"
/db_xref="GI:15155749"
/translation="MNIAVSKPPLTIRLCGPRGACGADRAIQIIVLAKAVFVY
RHEIVNRVYVELEKGAIVEELHEIPAEHREBPVYESAGSGVEDQOANLF
VIDACPLVSKHKOMRHOGRHVHVLIGHAGHPEVIGTMOLPEGVSVLVEYEDA
GYEYPRDRENLGFVOTITLSDPTAGVITARLDERPPIAIOAPRADSTCATTPKODAVK
QAPGCDLFIIVGAPNSSSKRLVEVALAGAKHSVLVORASEIDMNEIGDIRTVGLS
AGASAEVIVDEIIAEAFKARFDTLIDLAVTEETHFLVRELRSIELTDDMAFVNG
NASNALPKRATAGI"
9312..10280
/gene="AGR_C_1416"
9312..10280
/gene="AGR_C_1416"
/translation="HOMOSERINE KINASE (HK)"
/codon_start=1
/transl_table=11
/product="AGR_C_1416p"
/protein_id="AAK86584.1"
/db_xref="GI:15155750"
/translation="MAVYTDITDELDNELNLTQYDVGSLTSYKGIAGEVENSFLLHTT
KPLILTYLEKVERKNDLPFLGMQHLAAKGLISPLPRKDGELIELSGRPAALI
SFLGMLWLRKPEAKHCREVGAALAMHLASEGFEIKRPALISVDGMKYLMDSEERAD
EYKGIREEIRPEIDYLAHWPRLDPAVITHADLPDQNVFELGDELISGIDYFACND
LLAYDYSICLNAMCPENKDSAYVNTKGALEBEGYSVRPSEELPILSLSGSLRFR
FLRLYDWLTTPAGALVYKPDLETILKRLRHRTIANVAEYGLAGE"
10277..10717
/gene="AGR_C_1417"
10277..10717
/gene="AGR_C_1417"
/translation="Ribonuclease H (E.C.3.1.26.4) Mutant with His 62"

```

```

Replaced By Pro (H6ZP)
/codon_start=1
/transl_table=11
/product="AGR_C_1417p"
/protein_id="AAK86585.1"
/db_xref="GI:15155751"
/translation="MKHVDIFTDGASGNPSPGMAVLYRGETEKEISGGEADPTNN
RMELIAISALNALSPCEVDLYTDSYVYDGIITKWI FPKKKKMGKTAADNKPVKNNEL
MOALEAEOERHKVTLHWKRGHAGHPERADDELARKGMEPFKR"
BASE COUNT 2021 a 3194 c 3242 g 2378 t
ORIGIN

```

```

Query Match 44.2%; Score 740.4; DB 1; Length 10835;
Best Local Similarity 68.0%; Pred. No. 2.6e-145;
Matches 1109; Conservative 0; Mismatches 511; Indels 12; Gaps 5;

```

```

QY 13 gccattcaagccatgaccacatgagaagaagcctcttcacgcctgattatg 72
DB 2025 GACGATCATTCGATGATGATCCCATAGCGCGGCTTTTCGCGCTGCTGTC 2084
QY 73 accaaccaagaagatgctgctgctgctgctgctgctgctgctgctgctgct 132
DB 2085 ACCAACCAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2144
QY 133 tccgctccttcacgcctgctgctgctgctgctgctgctgctgctgctgct 192
DB 2145 GCGCGCGGCTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2204
QY 133 tgccttgaagcgcagctcgtatgctgctgctgctgctgctgctgctgctgct 248
DB 2205 CACGCTCGGCTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 2264
QY 249 acacctgtgagcagctgctgctgctgctgctgctgctgctgctgctgctgct 308
DB 2265 -CACATGTTCAACGCTTTCACGACGACGACGACGCTGATGATGATGATGAT 2323
QY 309 cccgcagcttgcgtgctgctgctgctgctgctgctgctgctgctgctgctgct 368
DB 2324 GCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2383
QY 369 tatgctccttcacgcctgctgctgctgctgctgctgctgctgctgctgctgct 428
DB 2384 CATGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCGCTTTCGCT 2442
QY 429 gggcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 488
DB 2443 GCTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2502
QY 489 ggtctgtacccgcgcctgctgctgctgctgctgctgctgctgctgctgctgct 545
DB 2503 GACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2562
QY 546 ttgcggttcaactgctgctgctgctgctgctgctgctgctgctgctgctgctgct 605
DB 2563 CTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2622
QY 606 cttcttgaacatgctgctgctgctgctgctgctgctgctgctgctgctgctgct 665
DB 2623 CATGCTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2682
QY 666 gattcttatacagcgtcgtgctgctgctgctgctgctgctgctgctgctgctgct 725
DB 2683 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2742
QY 726 catgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 785
DB 2743 CATGCTTCTGACCGACCGTACCTTGGCAGCGGCTTCTTCTGCGCGAGGCGGCGG 2802
QY 786 ccgcatctgtacacacatcctgctgctgctgctgctgctgctgctgctgctgctgct 845
DB 2803 TCCGATCCCTTACGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2862

```



```

/gene="mlr7090"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53264.1"
/db_xref="GI:1402668"
/translation="MAGKAKTDSPLVAARMSDIAYERILEGLFEDRRVPAGATISON
ELSATVDPAVAPLRDALRVLAEGLITTHPRSGIOFVPPGLELTRATYQFISYERAA
VRVPAEGDEAVMANLEIRHSRLRVRERDGLKQLEEMALFEGELHGAVALRNP
LIDSTYRRMHNLRLLDRKLTTPIMVRLKEHLDILEACGTNRADAAEKLAHQ
AAMQRSIGLV"
/gene=".5922
/4684..5922
/gene="mlr7091"
/4684..5922
/gene="mlr7091"
/codon_start=1
/transl_table=11
/product="sugar ABC transporter, periplasmic binding
protein"
/protein_id="BAB53265.1"
/db_xref="GI:1402669"
/translation="MQLKROFLMSALAMAGHTRCARASATSYHHPTSOSEMAS
LKITGLFOANADVVTOENIPNSEIYAKYSSAVLAGRPDTAVIHERFADLTAE
GLIDLDVNGWKRKANPENRMWGLSSNGAIYAVPAFVDMWYRADYFEEGLAG
PRNDEFELACRLTDPKSKRYAFAGMAGAGGKYVIDVMEATGSPYKDGOPADIR
AAIETATFYSDDLFLKEKVPSPANDSYROIEMEGFTQOTAMVHHTGSINELI
KOGQFSTAPMAGPKAHAVARYAVANGIMKEDNTEAMRWVSEMGOTDALTLLEGT
GTEPSAALODPRTGPNPIYAAVETLEFGRLPNNFCAAGWSENVNPAFOVLIG
OSTPQAVDRMIOGLEAALR"
/gene=".6932
/5979..6932
/gene="mlr7093"
/5979..6932
/gene="mlr7093"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="BAB53266.1"
/db_xref="GI:1402670"
/translation="MLSHRGSMTSSSGYRPLRALGESETRYMYMLLLESLVWILL
VIAPYLVGMATSFREMRLTRGSGWGEFKHYAAMSDPYEWALKNTYVAVVAVV
LELAGLLAATNLNRDLCGTKEFVILILPYLPVAVAGHMMALLDRLGIVINDLIY
RAGVLOTYKAMPADNPNTALATVLEAMHGFPPFALLLAGLKSIPEDLYKAAAVDA
GVSQDLRLTTPMLKTVIAAAVILRVIGLVNSPDLILLTGSGEDATQVLSYAFOT
AKENFEGYASALSVMFVILMIFATAYIRLSRIRYKGGDD"
/gene=".7770
/6925..7770
/gene="mlr7094"
/6925..7770
/gene="mlr7094"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="BAB53267.1"
/db_xref="GI:1402671"
/translation="MIESRRRRVIANIATYVLTLSKFCGPMHMFJLSKHEAD
IYVOTMOYIPRRVTDNVVAIWTOSGPEPLVANSVVFVLTIVATGTLAAVAFSR
FAFRGTQMLGVLYIVRMFPAVMMIIPLVAVNGVLDSRGLAATSPLLPLEVW
MLKGFDSAPKELESARIIDGSTRGAMRIYMPIARGLVASSEVFIATIAAMNEFLA
LMITGGGSRTPVGLQLMVGEFOLPMGVLAAGLILSTILPVLPAIVQRTWVQCLTA
GAIKG"
/gene=".8860
/7781..8860
/gene="mlr7095"
/7781..8860
/gene="mlr7095"
/codon_start=1
/transl_table=11
/product="sugar ABC transporter, ATP-binding protein"
/protein_id="BAB53268.1"
/db_xref="GI:1402672"
/translation="MATLISIKDVRRKSYGSAVQVLAHGVDELDEGFLVLLGPGSGCKST
LNIITAGLDTSQGDILLIGRSVVDLAPKDRINIAMFOSYALPYTMSVERNIGFLEM
RGVDADRRRAVDADARLLOISHLDRRANISGGOBROWANGRAIVRPELLEDEP
LSMLDKLVEKRTETIKRHERIGTSIYVTTDOIEAMTLAKVAVMMGCHTQOPADP
RTYERPAVFAVSFTGSPAMNLFGRVLSDATGVFAASGVSAIOLACKRTVDSERP

```

```

VIIGRIPEELKAVGRDEASVGVIDVEFTGPETWTVQVGDQSIILARLPFRAGRN
EPIFLAVDPASVNLFDPPGSEHRIDI"
/gene=".9082
/8654..9082
/gene="msr7097"
/8654..9082
/gene="msr7097"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53269.1"
/db_xref="GI:1402673"
/translation="MDAKALLAQRGCHRPMSRSHHRCRRDRRRLQRLMAVAEAY
ACGYRLXGEINQMIIPMWGRIAARQART"
/gene=".10692
/9079..10692
/gene="mlr7098"
/9079..10692
/gene="mlr7098"
/codon_start=1
/transl_table=11
/product="GMC-type oxidoreductase"
/protein_id="BAB53270.1"
/db_xref="GI:1402674"
/translation="MSPTSPFGDYVYVYGAGSAGCVLANRLSADPRNRVLLLEAGS
DRTHWVIDPIGTLFCMGNPTDMMKTEBPACINGRSLNTPRCKVIGCCSSVNGMYM
RGDAADYDGRQAGNMGWMDVLPYFRKSEDIHGKSHMHSGGEMVERORLSMPT
LDFAERDAEELGIPIRTGDFNTGNEGSGYEVNORLGVWNTSKAFRLRQIAORLNDV
VTGAEETRELEFEDRRVYGLLEFGDGRDHRVSAQOTILAGAINSPILBLSGVRDL
LSSIGVVRALPGVGENLDQHIQIRFVEVNAATLNERYHNLISRAAGVEYALAKR
SGPLSMASOLGIFARSADLATPDLIEHVQSTORLPPLRFPVAVVSCNLEPE
SRGVYHMASSDPREARTIRPNYSITACDGRSVAAVSRHVSRLMKARAVAFAPPEMLP
GTOTDSEDELIRKAGDIGTTFHFVGTCKMGSMAVVDRLRVHGLAQLRVYDASIM
PTIVSGNTSPVVMIEKAEMTILDPRR"
/gene=".11801
/10929..11801
/gene="mlr7099"
/10929..11801
/gene="mlr7099"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="BAB53271.1"

```

Query Match 43.0%; Score 719.6; DB 1; Length 346510;
Best Local Similarity 65.7%; Pred. No. 3.2e-141;
Matches 1097; Conservative 0; Mismatches 364; Indels 9; Gaps 3;

```

QY 1 atggcagaagcgcgcacccatcacgacgcatgaccacatgagaagcaaggctcttcacgagc 60
|||||
DB 337162 ATGGCAGAGCTGCGACGCTGACGACGAGCGCCACGACCAAGCCTTATCATGCGTGGTGCGC 337221
QY 61 tggctcaatgctgaccacacacaaagacatcggtctgctataccttctgagcgctggtgt 120
|||||
DB 337222 TGGGCTCTATTGACCAACACAGACATCGGACGCTCTACCTGATCTTCGCGATCATG 337281
QY 121 gttgttcattcttcgtctgctacacgctcacaagcgcttcgagatgagtcgagtc 180
|||||
DB 337282 GCGGCGATCATCGGCGCGCTGTGCGGTGCGCATCCGATCGATGAGATGTCGAAGACCTGCGC 337341
QY 181 gttcagtaacatgctcgtcgaagcgcaagtcgacgctgacgagtcgagtcgacgagaca 240
|||||
DB 337342 ATCCAGATCTTTCAGCGGCTGTGCGCGCAGATGCTATGCGCATGAACGATGAGCGCGCAT 337401
QY 241 gcgaacgagaca---ccgtgtgaagcgtcatggttacctacatggtatctgatagtcttc 297
|||||
DB 337402 GATGCGCGCAGACAGCATGTACATGCTTCGCGACCGCGCATGCGCTGATCATGATCTTC 337461
QY 298 ttgttggtatcccgccgctgtcgtggttctggttctgtaactatcgaatcgccgtcaaatc 357
|||||
DB 337462 TTCAATGATGATGCGCGGCTGTGATCGCGGCTTCGCCAACATGATGATGATGATGATC 337521
QY 358 ggcgctccgagatagtccttcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 417
|||||
DB 337522 GCGCGCGCGCGGACATGCGGTTCGCCGCGCATGAACACATCTCTCTGCTGCTGCGCGCGC 337581

```



```

/translaction="MATASANDHADDHHAHPTGKREASTNHRKIDTMYLVFA
IVAGIIGGMSGIRAEIMVPGIOVEPAHNYNEATFHCILIMEFVMPMTGEGN
WVPLITIGADMPAPRNNISITFMLIPAFALLISSMEGPGATGAGTWTATPIS
TSGHGPVADVSITSLIHLAGASSITGAINFTTTIFNMARPMTHKRLPFWSTLVTV
FLLLALPLVLAGAITMLITDRNFETGTAADGGPVLFQLEFMEFGHEVYLILPIS
FGMISHVSTESKPIFGYLGMAVMAVIGIGVMAHMYTVGMVSSAAYVVAAT
MIAVPTGVKFSVMITAMMGSIIEKFPMLFAIGIFLEFVGVTVGLAAGADRW
OETVYVIAHFVMTIAVFAIFSGMYWPMKSGMYNENYIKGLHFVMTIGANILF
PFOHFLGGMRRMVDYDPAFAGMNEISYSGAFITGSHVILPLYGLDAPARRQQA
DNPMGEATMLEMTISSPPPHQSTLTKVQ"
314.4303
/gene="cox3"
314.4303
/gene="cox3"
314.4303
/codon_start=1
/transl_table=11
/product="heme O synthase"
/protein_id="CAA61745.1"
/db_xref="GI:1177575"
/db_xref="SPTREMBL:O51322"
/translaction="MCAAHSCAHPRTKAGSTVSVDOKAVELPRISFAGVADYFALL
KPRMSLVITFTALVGLMTAPGHVHPVLGFIALLCIVACAGASGALNMLBEDIVLMS
RTANRPIDPRGRTTREGAMFGLTJSFVSMTDGLVNVAGGLAFITFEVYITMG
LKRRYQNVIVIGAGALPVPVMAAATGSLSEPLLFLLIFEMTPHFWALAFRS
DDYARAGVPMLEFVAVGADPATRLQILLYIVIVAAAPWPLGFEPMIVGWSLITGAG
MLVAIEVYRHRHRTGSOALRATRLPAFISIFLYLFAVLLLDVAKAVAPLIW"
BASE COUNT 749 a 1252 c 1415 g 964 t
ORIGIN

```

```

Query Match 42.4%; Score 709.2; DB 1; Length 4380;
Best Local Similarity 70.5%; Pred. No. 1e-138;
Matches 977; Conservative 0; Mismatches 403; Indels 6; Gaps 2;

```

```

QY 259 aaagtcagtgctccaccatggtatctcgatgagtctcttggtggatcccccagatg 318
1910 AACGTTTTCGACCAATCGCGCTGATCATCTTCTTCAATGGTGAATGCTCGGATG 1969
QY 319 ttcggtggttttgtaactatcgtatcgatcgatcgatcgatcgatcgatcgatcgatcg 378
1970 ATCGGGGGCTTGGCAACTGTTGTCGGCTGATGATCGCGCGCGCGCGCAATGCGCTTC 2029
QY 379 cgcgcatatgaacaactgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 438
2030 CCGCGGATGAAACAATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2089
QY 439 tgcgtgttcgacccggtgacggtcgaatcgatcgatcgatcgatcgatcgatcgatcg 498
2090 TCTCTGTTTCATGAGGGGTGACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2149
QY 499 cgcgcatgctgaaccgctg---aagctgtgctatcgtatcgatcgatcgatcgatcgat 555
2150 GCGCGGCTGTCAGAGCGGTGATCCGGGTCGGGCTGAGCTTCAAGCATTCGTGATGATC 2209
QY 556 caactgtcggtgtcctctctgtatcgaatggtggtggtggtggtggtggtggtggtggt 615
2210 CACCTTGGCGGGCGCTCTCTGATCTCGGGCGATCACTTCATCAACAGATCTTCAAC 2269
QY 616 atgcgccccccgcatgaagtcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 675
2270 ATGCGCGCGCTTGGCATGACCATGACCAAGATCCGCTGTCGTGTCGTGATCCTGGTTC 2329
QY 676 acggtgtgtgtatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 735
2330 ACCGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2389
QY 736 acggtgtgtgtatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 795
2390 ACCGACCGTAACTTGGCACACGCTTCTTGGCGCGCATGCGCGCGCATCCGCTGCTG 2449
QY 796 taccacaacacccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 855
2450 TTCCAGCATCTGTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2509

```

```

QY 856 ttcgcatcatcaagcactgtctgtctgacactctcgaagaaagccggtcttcggttactctg 915
2510 TTCCGACATGATCACCAACATGATCTCGACGTTCTCGAAGAACCGGATCTTCGGATATCTG 2569
QY 916 ccgattgtctatgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 975
2570 GGCATGCGCTTACGCGCATGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATG 2629
QY 976 atgtaacccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1035
2630 ATGTACACGCTGCGCATGCTGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCTG 2689
QY 1036 atcgcgtgtccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1095
2690 ATCGGCGTCCGCGACGCGGTGATGAGTGTCTGCGATGCGCATGCGCATGCGCATGCGCATG 2749
QY 1096 gttgagtcaaatcgcgcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1155
2750 ATCGAGTTCAAGACGCGCATGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCTG 2809
QY 1156 gttgtgaccggtatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1215
2810 GCGGTGACCGCGCTGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCTG 2869
QY 1216 taagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1275
2870 TACGTATCGCGCATGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCTG 2929
QY 1276 ggtatctacttctatcatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1335
2930 GCGTGTGCTACTGCTGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCTG 2989
QY 1336 ctgcactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1395
2990 CTGCATCTTCTGCGCATGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATG 3049
QY 1396 ggaagtcaggatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1455
3050 GGTCTGACGGGCTGATGATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCGCGCATGCTG 3109
QY 1456 aaagtcgtctactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1515
3110 GAGATCTGCTGCTGCGCATGCTGCTGCGCATGCGCGCATGCGCGCATGCGCGCATGCTGCTG 3169
QY 1516 ttgtctatacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1575
3170 CTGCTGACGCG---CTTCCGCGCGCAAGCAGCAGCGCGCTGACAAATCCGTGGGTGAAGGC 3226
QY 1576 gccgatacgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1635
3227 GCGACGACGCTGATGATGAGCCTGCTGCTGCGCATGCGCGCATGCGCGCATGCTGCTGCTG 3286
QY 1636 cccaag 1641
3287 CCGAAG 3292

```

```

RESULT 10
BYCORG 1830 bp DNA BCT 12-SEP-1993
LOCUS B.japonicum cox3 gene for cytochrome aa(3) subunit I (BC 1.9.3.1).
DEFINITION X54318
ACCESSION X54318.1 GI:39505
VERSION cox3 gene, cytochrome aa(3) subunit I.
KEYWORDS Bradyrhizobium japonicum.
SOURCE Bradyrhizobium japonicum.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Bradyrhizobium group; Bradyrhizobium.
REFERENCE 1 (bases 1 to 1830)
AUTHORS Gabel, C.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1990) Gabel C., Johns Hopkins University, Dept.

```



```

Db      1560 CTTATTGGGTGATCGATGCG---CTTTGGAAAGAGTGCCTGGCGGCGACAAATCCGTGG 1616
QY      1567 gggagattcgccgatcgcgtggaatggaacgcgcgcacccgcctccggccacacgltc 1626
Db      1617 ggTGGCGGCGGACGACGAGTGAGTGGACGCTGCCCTCCGCGCGCTTCATCATGATTG 1676
QY      1627 gaaacgctgccc 1638
Db      1677 GAAGTGTCTGCC 1688

RESULT 12
BUA242592
LOCUS   BUA242592
DEFINITION Bradyrhizobium japonicum coxB, coxA, coxE, coxF, coxG, coxC, sbhI
          genes: partial t1d gene and ORFs 133 and 177.
ACCESSION AJ242592.1 GI:6006409
VERSION   AJ242592.1 GI:6006409
KEYWORDS  assembly protein; coxA gene; coxB gene; coxC gene; coxE gene; coxF
          gene; coxG gene; cytochrome c oxidase subunit I; cytochrome c
          oxidase subunit II; cytochrome c oxidase subunit III; heme A
          synthase; Heme O synthase; hypothetical protein; ORF133; protoheme
          IX farnesyltransferase; sbhI gene; SUR1 homolog; t1d gene; t1d
          homolog.
SOURCE    Bradyrhizobium japonicum.
          Bradyrhizobium japonicum.
          Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
          Bradyrhizobium group; Bradyrhizobium.
REFERENCE 1 (bases 1 to 8121)
          Rossmann, R., Lofler, H., Rossi, P. and Hennecke, H.
          Factors involved in biogenesis of active cytochrome a3 encoded by
          the coxB/FCG gene cluster from Bradyrhizobium japonicum
          Unpublished
          2 (bases 1 to 8121)
          Rossmann, R.
          Direct Submission
          Submitted (20-MAY-1999) Rossmann R., Eidgenossische Technische
          Hochschule Zuerich, Mikrobiologisches Institut, Schmelzbergstrasse
          7, CH-8092 Zuerich, SWITZERLAND
FEATURES
  source   location/qualifiers
           1..8121
           /organism="Bradyrhizobium japonicum"
           /strain="110spc4"
           /db_xref="taxon:375"
           complement(1..440)
           /gene="t1d"
           complement(<1..440)
           /gene="t1d"
           /function="putative role in regulation of carbon storage"
           /codon_start=1
           /transl_table=1
           /product="t1d homolog"
           /protein_id="CAB56816.1"
           /db_xref="GI:6006410"
           /translation="MTNPDTTSLDRANLDRDQYRHRVARGLAGADDELFLFYSQTE
           ALMDNGRLKQATYDTSQSGFLRAVKDDAVGVAISSVSLPAILIRADAAVAAGGYS
           GSFSPPTHTVRLYGDNDPLDAPGFETKYLAEIDAYLRDKP"
           complement(542..1075)
           /note="ORF177"
           /codon_start=1
           /transl_table=1
           /product="hypothetical protein"
           /protein_id="CAB56817.1"
           /db_xref="GI:6006411"
           /translation="MGLSRMARPDSSLNALIGLAAAILLALPSLAHAGAVRSVHG
           DMOTRCDDPPGAGAECCALISVVAERNSAGLIVLYLKTADOKSLMRVVAPELGVLL
           PSGLGLKLDNDVGRASFVRCLENGCVAEVMDKLLGQLRTAKTATFIIFETPERGI
           GFPLSLNGLEGKDLK"
           1523..2362
           /gene="coxB"
           /gene="coxB"
           1523..1612
           /gene="coxB"
           /gene="coxB"

CDS
  gene     1523..2362
           /gene="coxB"
           /EC_number="1.9.3.1"
           /function="cytochrome c oxidation (respiration)"
           /codon_start=1
           /transl_table=1
           /evidence=experimental
           /product="cytochrome c oxidase subunit II"
           /protein_id="CAB56818.1"
           /db_xref="GI:6006412"
           /translation="MRMSKGRVGRHLLGLAVAGLTATGSAFAELGQAPAEWEPWLOO
           GSEPMNDIVWFHNFLEFVLLATLTLLVLLVAVVVRKFNARAPVSPRTHTNLTLEVA
           WTLVPLIIVGISVSPRLLELDVPKADLTIKATGKQWMSYAYPDDGKPEPDSIM
           AOKPRLGLGVNDPMVPPVKYRVQVGTADVTHARALPAFGYKIDAITGRINLTFK
           AKTKMFTYGCCESLCGRDIAEMPTAIRVEDKEFASWETAKKFASSGSGTGYASAG
           PTO"
           2451..4076
           /gene="coxA"
           2451..4076
           /gene="coxA"
           /EC_number="1.9.3.1"
           /function="cytochrome c oxidation (respiration)"
           /codon_start=1
           /transl_table=1
           /evidence=experimental
           /product="cytochrome c oxidase subunit I"
           /protein_id="CAB56819.1"
           /db_xref="GI:6006413"
           /translation="MATSAAHGDAODGHDEHAHPTGMRRYVSTNHKIDIGTWLLI
           FAVIAGVIGAMSLIATRAELMYPGVOIFETHTYVNFVSHGLIMFEFWMAMGIF
           GNEVPMICAPMAPPRMNNISFWLLPSGILLMTSTVEEGEPGNGAGAGTWYVP
           LSSGHPGAPVDPAITSLHAGASSITGAINPTTFNNRARGMTLHKRPLEWSTLV
           TVLLLSLPLVLAGAITMLTDNRETTFPAPDGGDPVLFQHLFEFHPVYLLIL
           PFGMISQVISTFSTRKPVFGYLGAAYVAIGIGVVAHHMYTVGMSASQAQFVA
           ATWVIAVPTGVKIFSWIATMGSGIEFRAPMTWAVGIFLFTGVGVLANAGVDR
           VLOETVYVVAHFHYVLSLGAFAIFAGWYVFPKMTGYVNETLAKAHFWVFEGNL
           VFEPQHLGLSGMPPRYVDYDDAFAGMWLVSSVSGYISGFGVLIPLCYIDAFPAKKV
           AGDNPAGAGTLEMTLPPSPPHQEVLPVNG"
           4137..5081
           /gene="coxE"
           4137..5081
           /gene="coxE"
           /function="protoheme IX farnesyltransferase"
           /codon_start=1
           /transl_table=1
           /product="putative heme O synthase"
           /protein_id="CAB56820.1"
           /db_xref="GI:6006414"
           /translation="MSVLQDNADVINDPRISEAEVGDYIALKPRWMSLYIFTALYGA
           MFGHPVLAITSILCTIAGAGSALMALEGGIDDAKMSRTANRPIPRGRTPEEA
           MPFGMTLAFESPWMTGIIYNTAGALLATFTFFVYIYTMKLRKTAONIYVIGAGA
           LPVYVAMAVTGVVDPEELLRLAIITFTPTPPHMLALFRSDYTRAGIPMLPNVAGP
           DATRLQILFTTVLVAVANAPMALGTFDVAIVCVSITLIGAGLVLAIVNMYRRSOS
           LRATRKLFASFSLIYLFALFTLLAEVFRALPVMGA"
           5078..5251
           /gene="coxF"
           5078..5251
           /gene="coxF"
           /codon_start=1
           /transl_table=1
           /product="hypothetical protein"
           /protein_id="CAB56821.1"
           /db_xref="GI:6006415"
           /translation="MTAAVADKPEPGICVITLTKQKSRKRSRISALALGVLVLFPAV
           TWKGRVAVLVRPM"
           5257..5901
           /gene="coxG"
           5257..5901
           /gene="coxG"
           /function="biogenesis of cytochrome c oxidase"
           /codon_start=1
           /transl_table=1
           /evidence=experimental

```


QY	1567	gacgaattcgcgaatcagcttgaagacgcctgcataccagccctcgagccacagctc	1626
Db	3993	GGTGGCGGGCAGCACACGCTGGAGTGGACGCTGCCCTCGCGCGCCCTTCATCATGTC	4052
QY	1627	gaacagctgcgc	1638
Db	4053	GAAGTGTGTCGCC	4064
RESULT	13		
LOCUS	BUJ33883/c		
DEFINITION	BUJ33883	31495 bp	DNA
			ECT
			10-AUG-2000
			Bradyrhizobium japonicum putative epoxide hydrolase Ephb (ephb), putative stress-induced protein Ohr (ohr), putative transmembrane transcriptional regulator protein Tirr (tirr), putative two-component system transcriptional response regulator Tcsr (tcsr), integral inner membrane metabolite transport protein MtbA (mtbA), orf157, and orf235 genes, complete cds; Fo ATPase operon, complete sequence; probable acetyltransferase Pact (pact), mitochondrial processing peptidase-like protein Mpp (mpp), threonine synthase Thrc (thrc), two-component system transcriptional regulator Elmr (elmr), putative two-component system sensor histidine kinase Elms (elms), SUR1-like protein (shd1), and orf133 genes, complete cds; cox operon, complete sequence; orf177, Tld (tld), signal peptidase Sips (sips), and orf375 genes, complete cds; and putative glutathione synthetase Gsha (gsha) gene, partial cds.
ACCESSION	U33883		
VERSION	U33883.2		
KEYWORDS	GI:8708897		
SOURCE	Bradyrhizobium japonicum.		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Bradyrhizobium group; Bradyrhizobium.		
REFERENCE	Muller,P., Ahrens,K., Keller,T. and Klauke,A.		
AUTHORS	1 (bases 1 to 31495)		
TITLE	A Tnpba insertion within the Bradyrhizobium japonicum sips gene, homologous to prokaryotic signal peptidases, results in extensive changes in the expression of PBW-specific nodulins of infected soybean (Glycine max) cells		
JOURNAL	Mol. Microbiol.	18 (5),	831-840 (1995)
MEDLINE	96422470		
PUBMED	8825087		
REFERENCE	2 (bases 1 to 31495)		
AUTHORS	Mueller,P.		
TITLE	Extended sequencing of a DNA fragment of B. japonicum adjacent to the cox operon		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 31495)		
AUTHORS	Mueller,P.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-AUG-1995) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Fritsch-Str., Marburg 35032, Germany		
REFERENCE	4 (bases 1 to 31495)		
AUTHORS	Mueller,P.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2000) FB Biologie, Zellbiologie und Angewandte Botanik, Philipps Universitaet Marburg, Karl-von-Fritsch-Str., Marburg 35032, Germany		
REMARK	Sequence update by submitter		
COMMENT	On Jun 26, 2000 this sequence version replaced gi:1200339.		
FEATURES	Location/Qualifiers		
source	1..31495		
	/organism="Bradyrhizobium japonicum"		
	/strain="USDA110spc4"		
	/db_xref="taxon:375"		
	/clone="C27"		
gene	complement(7..909)		
	/gene="epbA"		
	complement(7..909)		
	/gene="epbB"		
CDS	/codon_start=1		

[illegible]

Oy	1387	cacttccctggagacgcaaggatattgcacgcaccgtttcatcagtaccttcgaacgcccttcg	1446
Dd	23234	CACCTTCCTGGCCCTGTGCGGCATGCCGCCCTTACTGCTACCTCAGCAATCCGACGCGTTTGCC	23175
Oy	1447	ctgttgaaanaaacctctgcatatgttgcttcctcgaccctgcctgcctgatctgtcttc	1506
Dd	23174	GSGCTGAACCTGTCTCGTCGCGCTCFACATTTCGGCGCTTGCGCTCCTGATCTTC	23115
Oy	1507	accgtgacttttgtctatacgtctgtgttccttgccgcgcgcagagaacccttcgaacccgtg	1566
Dd	23114	CTCTATTGCGTAGTATCATTC---CTTTGGAAAAMAAAGCTGCCCGCTGGCGACATCCGTGG	23058
Oy	1567	ggcgaattccgcgatcacctgtaataaggaacgtgcacataacgcgcccccgcacaacgttc	1626
Dd	23057	GGTGGCCGGCGCACACCACGCTGAGTGGACGCTGCCCTTGCGCGCGCCCTTCATCATGTTTC	22998
Oy	1627	gaaacgcctgcacc 1638 Db 22997 GAATGCTGTGCC 22986	
RESULT 14			
RLOCXA			
DEFINITION	RLOCXA 1707 bp DNA	BCT	25-APR-1994
ACCESSION	X74341		
VERSION	X74341.1 GI:396698		
KEYWORDS	cytochrome c oxidase; cytochrome c oxidase subunit I.		
SOURCE	Rhizobium leguminosarum. Rhizobium leguminosarum Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.		
ORGANISM	1 (bases 1 to 1707)		
REFERENCE			
AUTHORS	Gabel,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JULY-1993) C. Gabel, The Johns Hopkins University,		
REDACTED	Dept. of Biology, Charles and 34th Sts., Baltimore MD 21218, USA		
REFERENCE	2 (bases 1 to 1707)		
AUTHORS	Gabel,C., Bittinger,M.A. and Maier,R.J.		
TITLE	Cytochrome a3 gene regulation in members of the family		
REDACTED	Rhizobiaceae: comparison of copper and oxygen effects in		
FEATURES	Bradyrhizobium japonicum and Rhizobium tropici Appl. Environ. Microbiol. 60 (1), 141-148 (1994) 94161496		
JOURNAL			
MEDIAE			
FEATURES			
SOURCE	Location/Qualifiers		
	1..1707		
	/organism="Rhizobium leguminosarum"		
	/strain="USDA 27"		
	/db_xref="taxon:384"		
gene	91..1707		
	/gene="coxA"		
	91..1707		
	/EC_number="1.9.3.1"		
	/codon_start=1		
	/transl_table=11		
	/product="Cytochrome-c oxidase, subunit I"		
	/protein_id="CAA52388.1"		
	/db_xref="GI:396698"		
	/db_xref="Swiss-Protein:P00885"		
	/translation="MATSAAAHGHAEHDHGDHDEHAHPTRNGRRSRNHKDTGLYLIPAI IAGVIGASLAIIRALMTPGVEVFENHTHLNFTVSRYGYIMFFMYPMPIMGFGMW FLPLMATPMPMRPNRNISFWLLPSAFGLLNSTFGEPEGANGACAQMVMYPVLSS SGHFGRVADLAIFSLMLTAGASSILGANFTITTLNMARPWLTLMKPLFMWSVLTFAF LLLSLPDLVGAILTMLTDNRNFRTTFEPARGGGDPLLQHLFMFGHEVYILTPEF GMIIISTFSRKRFEGYIGMYAMAIGLEFYVAHHNYIVOMDIDTEAYPVSATM IAWPTGIKIEMSWIAMGGSIEPATPMNLAIPLFTYGGVYAVLANASLDRLVH DYTYVVAPHVVISLGALIFAIFAGWMYPRKMSQMYNPFLAEHPFLITGVNIJEF BEHYHGSHPPRIIDPDPAFCGNLYSSITSGFSVLLFTCYVDAPAKNVPSD NPWGAGVTILEWLTPSPPEHVLELRVE"		
CDS			
BASE COUNT	277 a 537 c 479 g 414 t		
ORIGIN			

Query Match Similarity		39.13	Score 654	DB 1	Length 1707	
Best Local Similarity		67.77	Pred. No. 4	4e-127		
Matches 947		Conservative	0	Mismatches 445	Indels 6	Gaps
Qy	244	aacgcgaacccctgtagaacgtcatggttacctacacatggtatctctgatatggtctctgtg	303			
Db	301	AACGACCCATCTGTACAAAGTATWGTGGTATACCGGAGGGGTTATCATGATCTTCTCATG	360			
Qy	304	ggtatcccccgatattgttctgtgttctttgttaactatctgatactgctgcgtctcaatctgcgt	363			
Db	361	GTCAATGCCGGCTATGATATCGGGGCTTGGGACATGTTCTCGCCCTGTATGATATGGCGCC	420			
Qy	364	ccgcgaatacgtgcttcccgcgcgtatgaacaacctctgtctctgctgtgtctcaatctccgttaac	423			
Db	421	CCCGATATGGCGCTTCCCGCGGATTAACAATATCTGTTCTGCGCTCTCCCGCGCTTTCGTT	480			
Qy	424	gcgatbaggcgtgctctgcgtgtctgcacacgcgcgcgtgacagctgaagcttggtctgcgcgt	483			
Db	481	GGTCTGCTGCTCATATGATGACCTTGTGTGAGAGGGGAACCGGCTGCCAATAGGCGCGGTGG	540			
Qy	484	ggttgggtctgttaaccgcgcgtctg---tgaaaccgcgaagctgtgcattctgattgacctc	540			
Db	541	GCTTGACAGATGATATGTGTCGGCTGTCAATCAAGGCCATTCGGGCCCGCGGTTATATCG	600			
Qy	541	gcgaatttcgcgcgtctcaactctgacgcgtgctctcctgatactgagcgcgaatacaataatc	600			
Db	601	GCAATCTTTTGGCGATATTTGGGGGGCGCTGTGTCATCTCGTGGCGATCAATCTTCATC	660			
Qy	601	acgacctctctgaaacatgacgcgcgcccgacatagcgcgtgcacaaagtgcgttgttctcg	660			
Db	661	ACCAAGATTCCTGAACATATCGCGGCTCCCGGCTATACCTGCACAAAGATGCGCGTGTATGCC	720			
Qy	661	tgtctgaccttatcaacgcgccttgagctgcatactcgtctgcgcgtcgcggtttctgctgtgca	720			
Db	721	TGTCGTGTTTGAATACCGCATTCCTGCTCTTCTGTCACTCCCGGTTTCGCGCGAGCGC	780			
Qy	721	atcaacaatgctgctgacacgcgcgttaactctgcgcgcgcacttctcaactctgctgcgcgc	780			
Db	781	ATCACACATGCTGTGACCGACCGCAATTTGGGACACCTTTCGGCGCGGAGAGGCGGC	840			
Qy	781	ggttaaccggaattctgtctcaacaaacatctcgtgtgttcttttgggcacccgcgaagtgaacatc	840			
Db	841	GGTATCCGCGTCTCTACACAGCACTTCTTGTGTTCGGCATTCGGGAATCTCAATTT	900			
Qy	841	atcaattctgcgcgcgtcttgatcatcatgaacgaagctgtctgacactctcgcgaanaagcgc	900			
Db	901	CTGATMTCGTCGCCGGTTTCGGCATGATCAGCAATCATCTCAGACTTCTCGCGGCAAGCG	960			
Qy	901	gtctctggttlaactctgcgcgaatggtctatgcaatgagtggaactcgcgtgttctctgagctttgc	960			
Db	961	GTAATTCCGCTACATCGGCATGATATACGCTATMGCTATGAGGGGGCGCTCGGGTTTGTG	1020			
Qy	961	gtctgggcgcacaacatgataacacggtttgagtctgtctcaacgaacatctcaactaatg	1020			
Db	1021	GTCGCGGCGCACCAATGTAATGTCGCGCATGAGATCTGGATACCGAAGCGTATTTGCTC	1080			
Qy	1021	ctggaccacaatgtagatcgcgcgtgacgcgcgcgcgaatgaatctctctctgtagatcgcacgc	1080			
Db	1081	TGCGGCGACATGATCATWCTGTGTGCGACCGGATCAAGATCTTCATMGATGATTGCGAGC	1140			
Qy	1081	atgtcgggggcgcgtcgtgtgtggtttcaaatctgcgcgaatgctcttgaggcttttgctttatgttc	1140			
Db	1141	ATGTGGGGGTGTACATCGAATTCGCCACCCCATCTGTGGGCGCTGTGCGTTTATCTTC	1200			
Qy	1141	ctgtgtaacgcgtgtagtgtgtagaacgcgtatctgctgtccgcgaacgcgcgtctgtagacgttca	1200			
Db	1201	CTGTTCACGGTGTGGCGCGCTTACCGGGGTGTCTCTCGCTAACAGCGAGAGCTGATGGGGTC	1260			
Qy	1201	tatcaagaacacttatctgaagtgtgcgcgaactctccatctatgtagatgcgtctggtgtgacatc	1260			
Db	1261	CTCCAGATATCTTACATGCTGTGTGCTCATTTTCCATCTAGCTCTCTGCTGTGGGGCTATTT	1320			

PTFARKPLFGYTSWVYATASIAFLSEFVMAHMFVVGIPYIGELFMVATMLIAVPLG
 VKFVNWTTWMEGSLTFETPMLFAVAFVITLFTGGESGLMATAIPAFQVHDYFVA
 HFHYLVPGATFGIPASAYVYLPMWTHMDTELCKLHFMKSFIGNMLAFPMHFVGL
 AGMRRLPDVILQFADFNWSSIGAFMEGTTQLLFLFVILKIRGKPADAKPMDGAE
 GLEWISPSAPYHFSFPEVK"
 6463..7017
 /gene="PA0107"
 /gene="PA0107"
 6463..7017
 /transl_table=1
 /codon_start=1
 /product="conserved hypothetical protein"
 /protein_id="AAG03497.1"
 /db_xref="GI:9945934"
 /translation="MSDAKVDRLVGRLLLVYVLMFAGFALVPLDYWCALGNG
 KTASATSGEODVGRVYVQETSTNNIMWFERGAGOLVHPAVQMVYAN
 PSDKPMYQAIPISIAPEAAAYPHKTCFCFTQDVLOPGESEIMPVAFIVDRDLKDV
 RHVVLATLFDITARKPPVAGR"
 7028..7915
 /gene="coi11"
 /note="PA0108"
 7028..7915
 /gene="coi11"
 /codon_start=1
 /transl_table=1
 /product="cytochrome c oxidase, subunit III"
 /protein_id="AAG03498.1"
 /db_xref="GI:9945935"
 /translation="MASHEHYVYPAQSKWPIIASIGLTVTFGLTWENDLQAGHES
 HGVPIFVVGSLIAYMLEFGVNYIRSRAGLYSAQMDRSFNGMSFIFISEVWFPA
 FFGALFYVRHFRAGWLGEGAGVAHMLMFOYSWLLQTPPKLEPPSAVIEPMK
 LPLNTLLVTSFTVFAHNAALKKRGKPLKMLATLVLTGLTAPLLOAEVYHAYN
 ELGLTLAGIYGSFEMLTGPHGNAVTLGALLIIGIMLIRLGRHFDAEHNGFEAASW
 YMHVVDVWVGIFEFYVIT"
 complement(7931..8140)
 /gene="PA0109"
 complement(7931..8140)
 /gene="PA0109"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAG03499.1"
 /db_xref="GI:9945936"
 /translation="MLKVALVLLLATLVLSFGLFELVKDQHGSRVNSLTVRVVL
 AAATLVVAMGFYSGELINSHAPMHF"
 8156..8950
 /gene="PA0110"
 8156..8950
 /gene="PA0110"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAG03500.1"
 /db_xref="GI:9945937"
 /translation="WGFQWYVSQLDEAGNDRSRMGAFRGLFTLVVLLPVLL
 WLCTWLOQRADERKALLASYEARGAEPVSPQLEGLRPAYVRVLLHGFEDRHLL
 LDNRRLNGAGVEVLQPFYDOASGLWMLPLVRCVAVWTDERSPTLLEPDRVLLDWT
 YLPFPGHLADAPAGMPLVLTQDLPALMOAGFAGAGLPWEIRLEPQDASFTDMPPL
 VSMPEHHTGYAVOWFALATALLATLVLYLGVRRAREKHNHSRSDA"
 8925..9503
 /gene="PA0111"
 8925..9503
 /gene="PA0111"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAG03501.1"
 /db_xref="GI:9945938"
 /translation="MSLAIPMERPRRARGRLQILAIGLVGPMLLASAMYKNFMV
 PGRSYSGALIGNGOTPADLGVGRSGEGMOLLVAPGCGECCOOLVYLKQINIA
 LGREASRAGHALAAGELPADPANTVRODPRIORAGDLPALYDKRGGGERPLWVVD
 PHGNIVRYDAKNGKKLLKDVQLLKLSHIG"
 9568..10641
 gene

CDS
 /gene="PA0112"
 9568..10641
 /gene="PA0112"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAG03502.1"
 /db_xref="GI:9945939"
 /translation="MSGATRKPGFTVAVALVLLAAVVLGAVTRLTHAGLCCPDMPG
 CYGFVHPLSEAOIAHAELEHPPDAPVAAKQKNGNEMHRYFAGALGLILGLALHLVR
 RGRDGOPLKPLILLAVVTAOAQFGWMTLKLMPQVYVHALLGGFTLLALLFLALR
 LSRFARARYPATRILAGLALLIVTQIALGWSVSNVAAYACIDLPICHEWPMRM

Query Match 30.8%; Score 515.6; DB 1; Length 12829;
 Best Local Similarity 63.0%; Pred. No. 3.2e-98;
 Matches 854; Conservative 0; Mismatches 484; Indels 18; Gaps 3;

QY 286 ctatgatgtctcttcttgaggtalcccgcatctgtcggtgttcttgtaactatctgatg 345
 Db 5106 CTGATCATGGCTCTTGCGCGCGGTATGCCGCTCTGTCGCCCACTGATGATC 5165

QY 346 ccgctgcaaatcgcgctccggaatgacctcccgatgacaacatgctctctg 405
 Db 5166 CCGCTGATGATCGCGCGCGGACATGCGCCCTCGCGGATGACAACTTCACCTTCTG 5225

QY 406 ctgttcattgcggtlacgcgcgatggcggtgcttcgctgttcgacccggcggtgaagat 465
 Db 5226 CTGCTGCGCGCGCGCTTGCGCGCTGTGTCAGCACCTTGTTCATGCCCGCGG----- 5278

QY 466 cagctgggttcggcgctgtgtgtgtctctgacccgctgcgacccgggaagctgac 525
 Db 5279 -----CGGCCCAACTTGCCTGACCTTATGCGCGCGCTGCGACCACTTGCGCCGCG 5333

QY 526 taltcgatgacctcggaatttcgacgttctgacgttgcgggttcctcctcgatcatgagc 585
 Db 5334 CACAGCCTGACCTTTCATCATCTTCGCATCCACCTGACCGCGGATCAGTGCATGAGCG 5393

QY 586 gcatcaagaatgacacgtcttgaacatgacgaccccgacgtacgtgcgaca 645
 Db 5394 GCGATCAGCTGATCGGCACCATCTCACTCGCGCGCGCGCGGATGACCCCTGATGAAG 5453

QY 646 gtgcgctgtctcgtgctgacatcatatcaacgcttgcgtacccgtgcgcgtgcg 705
 Db 5454 ATGCGGCTGTTCTGTCGAGACCTGCGTGCATGACCGGCTTCTTTGATCCGGGATGCGCT 5513

QY 706 gtcttgatctggtgacatcacatgctgacgacgacgacgacgacgacgacgacgac 765
 Db 5514 GTCTGCGCGCGCGTGTGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 5573

QY 766 aatcctgtctgacgacgtgacacgacatctgtgacacacacatcctggtcttctggac 825
 Db 5574 AGCG 5633

QY 826 ccggaagtgtacatcatctcgtcccgcttgcgtacatcatcagacatctgctgtgac 885
 Db 5634 CCGAGAGTGAACATCATGATCCTGCGCGCGCTTGCGGCGGTACAGTGCATTCGCGAC 5693

QY 886 ttctgaaaaagcggctctcgttaccctgcgacgtgctatgacatggttgcacatcgat 945
 Db 5694 TTGCGCGCGCAAGCGCGCTGTTGCGCTACACCTCGATGCTACGCCACCGCGACGCGCG 5753

QY 946 gtcttgagctctgtctgctggcgacacacatctgacacatctgctgctgacgacg 1005
 Db 5754 TTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5813

QY 1006 caatcctactcatctgctggcaccatggtgacggtgacgacgacgacgacgacgacgac 1065
 Db 5814 GAAGCTGTTCTTCAATGACCAACCATGATGATGATGATGATGATGATGATGATGATGATG 5873

QY 1066 tctgtgatcgcaacgctgtggtggcgctgcgtgtgagttcaatcgccgacgctcgtggcc 1125
 Db 5874 ACTGGGTACCAACCATGTTGAGGAGGCTGCTGACCTTGGAGACCGCGCATGCTGTTGCGCC 5933

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 08:10:54 ; Search time 228.32 Seconds
(without alignments)
6285.750 Million cell updates/sec

Title: US-09-712-768-1
Perfect score: 1674
Sequence: 1 atggcagacgcccacatca.....acaagcatccctgcacata 1674

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SID88/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SID88/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SID88/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SID88/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SID88/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SID88/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SID88/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SID88/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SID88/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SID88/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SID88/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SID88/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SID88/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SID88/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SID88/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SID88/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SID88/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SID88/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SID88/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SID88/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SID88/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1674	100.0	1674	22	AAAG1489
2	397.2	23.7	1539	19	AAV3718
3	397.2	23.7	1693	19	AAV53011
4	397.2	23.7	1738	16	AA705905
5	397.2	23.7	16569	20	AA211492
6	397.2	23.7	16569	20	AA211493
7	397.2	23.7	16569	21	AA211491
8	397.2	23.7	16569	21	AA257203
9	395.2	23.6	6691	21	AA257204
10	384.4	23.0	1735	22	AA568879
11	296.6	17.7	1596	22	AAH67733

12	296.6	17.7	1743	22	AAH67734	C glutamylcyclo codin
13	296.6	17.7	349980	22	AAH68532	C glutamylcyclo codin
14	295	17.6	1779	22	AA771686	Corynebacterium gl
15	269.6	16.1	1992	22	AAH84533	E. coli growth and
16	223	13.3	1947	22	AAH52833	S. epidermidis ope
17	223	13.3	2397	22	AAH53978	S. epidermidis gen
18	223	13.3	3298	22	AAH54428	S. epidermidis gen
19	216.4	12.9	673	22	AAH77960	CDNA encoding huma
20	216.4	12.9	673	22	AAH77960	Colon tumour relat
21	211.4	10.6	4121	18	AAV74626	Staphylococcus aur
22	180.4	10.8	603	21	AAV78259	Human cancer assoc
23	179.6	10.7	972	22	AA71688	Corynebacterium gl
24	160	9.6	168	22	AA91493	G. oxydans cytochr
25	158.6	9.5	482	21	AA77948	CDNA encoding huma
26	158.6	9.5	482	22	AA128686	Colon tumour relat
27	149.4	8.9	347	20	AAV8616	EST clone CJS47.
28	147.4	8.8	3639	22	AAH54029	S. epidermidis gen
29	142	8.5	438	21	AAH16371	Human prostate can
30	141.2	8.4	3465	22	AAH54890	S. epidermidis gen
31	141	8.4	484	21	AAH16338	Human colon cancer
32	137.2	8.2	587	21	AAH18356	Lung cancer associ
33	129	7.7	735	22	AA71689	Corynebacterium gl
34	128.4	7.7	627	16	AAH19852	Human gene signalu
35	122	7.3	626	21	AAH16350	Human colon cancer
36	121	7.2	594	21	AA75690	Human ORF1245
37	117.6	7.0	367	22	AAH50706	Human tumour assoc
38	117.6	7.0	367	22	AAH50706	Human tumour assoc
39	111	6.6	310	22	AAH54615	S. epidermidis gen
40	108	6.5	277	20	AAH37772	Mouse cytochrome o
41	108	6.5	277	21	AAH37772	Mouse cytochrome o
42	106.4	6.4	277	20	AAH13992	SEQ ID NO: 19 for
43	106.4	6.4	277	20	AAH13992	Cytochrome c oxida
44	106.4	6.4	277	21	AAH39698	L7/L21 consensus n
45	106.4	6.4	277	21	AAH32247	Mouse body weight- Mouse cytochrome c

ALIGNMENTS

RESULT 1	AAA91489	standard: DNA: 1674 BP.
ID	AAA91489	
XX	AAA91489:	
AC	06-AUG-2001 (first entry)	
XX		
DE	G. oxydans cytochrome C oxidase (COI) coding sequence.	
XX		
KM	Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;	
KM	oxidative fermentation; electron transfer; respiratory chain; L-sorbose;	
KM	2-Keto-L-gluconic acid; 2KGA production; aldehyde production;	
KM	carboxylic acid production; ketone production; ds.	
XX		
OS	Glucanobacter oxydans.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1674
FT		/tag= a
XX		/product= "COI"
XX		
PN	EP1103603-A2.	
XX		
PD	30-MAY-2001.	
XX		
PF	14-NOV-2000; 2000EP-0124785.	
XX		
PR	17-NOV-1999; 99EP-0122842.	
XX		
PA	(HOFF) HOFFMANN LA ROCHE & CO AG F.	
XX		
PI	Asakura A, Hoshino T, Shinjoh M;	
XX		

DR WPI: 2001-357953/38.
DR P-PSDB: AAV97750.

PT New cytochrome c oxidase complex having cytochrome c oxidase activity
PT from Gluconobacter oxydans DSM 4025, useful in mediating electron
PT transfer in respiratory chain or producing 2-keto-L-gulononic acid from
PT L-sorbose or D-sorbitol

PS Claim 13, Page 16-20; 42pp; English.

XX This sequence encodes the gluconobacter oxydans cytochrome C oxidase
XX (COI) of the invention. The COI complex is useful in improving oxidative
XX fermentation and is an essential component mediating electron transfer
XX in the respiratory chain. The recombinant microorganism and the
XX cytochrome c oxidase may be used in the genetic preparation of a
XX recombinant COI complex and in the production of 2-keto-L-gulononic acid
XX (2KGa) from L-sorbose or D-sorbitol in a culture medium. The COI is also
XX useful as a terminal oxidase, oxidizing cytochrome c, an electron
XX acceptor from an enzyme belonging to dehydrogenase for the production of
XX aldehydes, carboxylic acids and ketones from alcohols and aldehydes,
XX especially the production of 2KGa from L-sorbose or D-sorbitol.

XX Sequence 1674 BP: 270 A; 493 C; 468 G; 443 T; 0 other:

Query Match 100.0%; Score 1674; DB 22; Length 1674;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcagacgcgcacatcaagcgcatgaaccatgaagaagcgttctacgcgc 60
DB 1 atggcagacgcgcacatcaagcgcatgaaccatgaagaagcgttctacgcgc 60
QY 61 tggttatgctgacgcaacaaagacatcggtctgtactactcttbaacgctgtgt 120
DB 61 tggttatgctgacgcaacaaagacatcggtctgtactactcttbaacgctgtgt 120
QY 121 gttggtttacatccgctcctgttaccgcttactatgacgcttgaagctgaatccgggt 180
DB 121 gttggtttacatccgctcctgttaccgcttactatgacgcttgaagctgaatccgggt 180
QY 181 gtccagtaactgtgcttgaagcgacgtctgacgagatgacgctcgaagacatgacg 240
DB 181 gtccagtaactgtgcttgaagcgacgtctgacgagatgacgctcgaagacatgacg 240
QY 241 ggcgaacgaacctgtggaacgcacgtatcaactacatgtaattctgatagtctctt 300
DB 241 ggcgaacgaacctgtggaacgcacgtatcaactacatgtaattctgatagtctctt 300
QY 301 gtgggatacccgcatgttctggtgtttgttaactatctgaatgcgcgtgcaaatccgc 360
DB 301 gtgggatacccgcatgttctggtgtttgttaactatctgaatgcgcgtgcaaatccgc 360
QY 361 gtcccgagataatgcttcccggtatgaacaacctgtctgttctggtgttctatgctgggt 420
DB 361 gtcccgagataatgcttcccggtatgaacaacctgtctgttctggtgttctatgctgggt 420
QY 421 accgcgagtgagcgctgctcgtcttgcacgacggcggtgaacggttaagctgagttcgggc 480
DB 421 accgcgagtgagcgctgctcgtcttgcacgacggcggtgaacggttaagctgagttcgggc 480
QY 481 gttggttgggttctgtaaccgcgctgtgcacccgcggaagcttggctatctgaatgacctc 540
DB 481 gttggttgggttctgtaaccgcgctgtgcacccgcggaagcttggctatctgaatgacctc 540
QY 541 ggcatttcgggttcaactgtctggtgtctctctgatacatggtggcgcatcaacatgac 600
DB 541 ggcatttcgggttcaactgtctggtgtctctctgatacatggtggcgcatcaacatgac 600
QY 601 acgaacctctgaaacatgacgccccgcgacgaagacgtgacacaaagtgcggttctctcg 660
DB 601 acgaacctctgaaacatgacgccccgcgacgaagacgtgacacaaagtgcggttctctcg 660

QY 661 tggatcatttatacagcgcttgatccctgctgacgctgacgcttctgctggtgtga 720
DB 661 tggatcatttatacagcgcttgatccctgctgacgctgacgcttctgctggtgtga 720
QY 721 atcaacatgctgctgacgacgacgtaacttggcgcagaccttctcaatctctgtcgggc 780
DB 721 atcaacatgctgctgacgacgacgtaacttggcgcagaccttctcaatctctgtcgggc 780
QY 781 ggtacccgattctgtataccacaacacccctgtgttcttggcaccggaagtgtacatc 840
DB 781 ggtacccgattctgtataccacaacacccctgtgttcttggcaccggaagtgtacatc 840
QY 841 atcattctgcgcgcttggatcatcaacgaatctgtgtacgcttctcgaataaagcgc 900
DB 841 atcattctgcgcgcttggatcatcaacgaatctgtgtacgcttctcgaataaagcgc 900
QY 901 gttcttctgctacccgctgacgataatgtaacatggttggaatcgtgttctggtgtgtgc 960
DB 901 gttcttctgctacccgctgacgataatgtaacatggttggaatcgtgttctggtgtgtgc 960
QY 961 gttcttctgctacccgctgacgataatgtaacatggttggaatcgtgttctggtgtgtgc 1020
DB 961 gttcttctgctacccgctgacgataatgtaacatggttggaatcgtgttctggtgtgtgc 1020
QY 1021 ctggcaccacatgtgtatcgcgtgacgacgcaatgaatctctcgtgtgatccgcacg 1080
DB 1021 ctggcaccacatgtgtatcgcgtgacgacgcaatgaatctctcgtgtgatccgcacg 1080
QY 1081 atgtgggacgctgctggtggtgaatgaatgaatgacgacgcttctggtgtgtgtgtgc 1140
DB 1081 atgtgggacgctgctggtggtgaatgaatgaatgacgacgcttctggtgtgtgtgtgc 1140
QY 1141 ctgttcaacgctggtgc 1200
DB 1141 ctgttcaacgctggtgc 1200
QY 1201 taccagacacatatacgtgc 1260
DB 1201 taccagacacatatacgtgc 1260
QY 1261 ttggtgacttgcgcggtatctacttacttacttacttacttacttacttacttacttactt 1320
DB 1261 ttggtgacttgcgcggtatctacttacttacttacttacttacttacttacttacttactt 1320
QY 1321 gaatggctgcaaaagctgacacttctgacacttcttcaatcgtgtcgaagctcaactcttc 1380
DB 1321 gaatggctgcaaaagctgacacttctgacacttcttcaatcgtgtcgaagctcaactcttc 1380
QY 1381 ccgcagacacttcttgcgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 1440
DB 1381 ccgcagacacttcttgcgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 1440
QY 1441 ttgcgcgtgtgaaacaaagcttgcgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgc 1500
DB 1441 ttgcgcgtgtgaaacaaagcttgcgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgc 1500
QY 1501 ttcttcatcgtgtatcttctgtatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgc 1560
DB 1501 ttcttcatcgtgtatcttctgtatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgc 1560
QY 1561 ccgttgggacgacatctgcgcgatacgtgtgaatggaacgacgtgcacataccgctccgacac 1620
DB 1561 ccgttgggacgacatctgcgcgatacgtgtgaatggaacgacgtgcacataccgctccgacac 1620
QY 1621 acgttcgaaacgctgcgaacgacgtgcgaacgacgtgcgaacgacgtgcgaacgacgtgcga 1674
DB 1621 acgttcgaaacgctgcgaacgacgtgcgaacgacgtgcgaacgacgtgcgaacgacgtgcga 1674

RESULT 2
AAV43718
ID AAV43718 standard: cDNA to mRNA; 1539 BP.
XX

AC AAV43718;
XX
DT 16-NOV-1998 (first entry)
XX
DE Cancer associated gene fragment 3.
XX
KM ds: cancer; PCR: Northern blotting; ribonuclease protection assay;
KM diagnosis; metastatic cancer.
OS Synthetic.
XX
PN WO9837187-A1.
XX
PD 27-AUG-1998.
XX
PF 18-FEB-1998; 98WO-JP00667.
XX
PR 21-FEB-1997; 97JP-0052508.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Asada K, Hino F, Kato I, Mukai H, Yoshikawa Y;
XX
DR WPI: 1998-467552/40.
XX
PT Detection of cancer cells in tissue samples - by changes in mRNA
PT expression compared to normal tissue of specific cancer-associated
PT gene sequences
XX
PS Claim 1; Page 43-44; 92pp; Japanese.
XX
CC The cancer associated gene fragments AAV43716-V43731 can be used to
CC detect cancer cells in tissue samples or biological fluids. They are
CC detected by monitoring the change in mRNA expression as compared to
CC normal tissue of one or more cancer-associated genes whose cDNA
CC stringently hybridises to the nucleic acid fragments. The change in
CC expression may be an increase or a decrease compared to normal tissue.
CC The mRNA expression may be determined by PCR, Northern blotting or
CC ribonuclease protection assay, or by determining the change in the amount
CC of protein encoded by the gene(s) as compared to normal tissue, for
CC example by using a labelled antibody recognising the protein. Detection
CC of cancer cells for cancer diagnosis, including detection of metastatic
CC cancer cells in tissues other than the primary tumour site.
XX
SQ Sequence 1539 BP; 417 A; 463 C; 249 G; 410 T; 0 other;

Query Match 23.7%; Score 397.2; DB 19; Length 1539;
Best Local Similarity 58.1%; Pred. No. 1.6e-96;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2

244 aacggacacctgtggaacgctcagtgttaacctacacatgatactgtatgtctgtg 303
148 aacgaccacatctacaacgcttagtcgtcaacgacatgcatgttaatactcttcata 207
304 ggtatcccgcactgttcgtgttttggtaactatctgatgcgctgcaaatcgcgct 363
208 gtaataaccatcataatcgaggccttggcaactgactagttccccaataatcgtgtgc 267
364 ccggatattgacctcccggtatgacaacacccgtctgtctgctgttcatccggtacc 423
268 ccggatatggtgtcccccgcataaacaacataaagctcttgcactcttaacctcttc 327
424 gcgatggcggtgtgctgctgttcgcacccggcggtgacgctgaactggttcggagcgt 483
328 ctacctcctgtctgcacatctgtctatgtagggccggagcaggaaacagttgacacgtc-- 384
484 ggttggtgtctgtacccgcgcgtgtgcacccggcaagctgagctatcgatggaacctgcg 543
385 -----taacctccttagcaggaaactctccacaccggagagcctcgtagaaccaacc 438
544 atttcgggttcaacttgcgtgtcctctcgatcataatgagcgatcaacatgatcacg 603
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

[illegible]

DT 11-JAN-1999 (first entry)
 XX Human cytochrome c oxidase subunit I gene COX 1.
 XX
 XX COX 1 gene; cytochrome c oxidase; Alzheimer's disease; diagnosis;
 KM mitochondrial DNA; ds.
 XX
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..71
 FT /*tag- a
 FT 72..1613
 FT /*tag- b
 FT /product= cytochrome c oxidase subunit I
 FT /note= "corresponds to nucleotides 5904-7445 of the
 FT published Cambridge sequence for normal
 FT human COX 1"
 FT 3'UTR 1614..1693
 FT /*tag- c
 XX
 PN W09838335-A1.
 XX
 PD 03-SEP-1998.
 XX
 PF 27-FEB-1998; 98WO-0503429.
 XX
 PR 28-FEB-1997; 97US-0810599.
 XX
 PI (MITO-) MITOKOR.
 XX
 PI Fahy ED, Ghosh S, Herrnsdorf C, Parker WD:
 XX WPL; 1998-481216/41.
 XX
 PT Detecting the presence or risk of Alzheimer's disease - by detecting
 PT mutations in the sequence of a mitochondrial cytochrome c oxidase
 PT gene in mitochondrial nucleic acid
 XX
 PS Disclosure; Fig 1; 125pp; English.
 XX
 CC This is the 5' upstream non-coding region, the complete coding
 CC region encoding mitochondrial cytochrome c oxidase (COX) subunit I
 CC and the 3' downstream non-coding region of the human COX 1 gene.
 CC COX, an important terminal component of the electron transport
 CC chain located in mitochondria, is composed of at least 13 subunits,
 CC 3 of which (COX 1, COX 2 and COX 3) are encoded by mitochondrial
 CC genes. The invention relates to methods of detecting genetic
 CC mutations in mitochondrial COX genes that segregate with
 CC Alzheimer's disease (AD), and methods for determining the amount of
 CC heteroplasmy of mitochondrial nucleic acid. The invention
 CC provides methods for detecting such mutations, as a diagnostic of
 CC AD, either before or after the onset of clinical symptoms.
 CC Mutations especially associated with AD are missense and silent
 CC mutations located at nucleotide sites 6221, 6242, 6266, 6299, 6366,
 CC 6383, 6410, 6452, 6483, 6512, 6542, 6569, 6641, 6935, 6938 and 7146
 CC of the COX 1 gene (numbering according to the published Cambridge
 CC sequence for normal human COX 1), nucleotide sites 7650, 7705, 7810,
 CC 7868, 7891, 7912 and 8021 of the COX 2 gene (numbering according to
 CC the Cambridge sequence for normal human COX 2) (see AAV53012) and
 CC combinations thereof. Probes (see AAV53013-30) and kits useful for
 CC detecting AD-associated mutations are claimed.
 XX
 XX Sequence 1693 BP; 464 A; 509 C; 273 G; 447 T; 0 other;
 XX
 Query Match 23.7%; Score 397.2; DB 19; Length 1693;
 Best Local Similarity 58.1%; Pred. No. 1,6e-96;
 Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;
 QY 244 aacgacacgtggaacgcatggtactaccatgattcttgatgagctcttg 303
 DB 219 aacgacacatctaacggtatcgtacagccatgacattgtgaaatcttccata 278

QY 304 ggtatccccgacattgtctggttttggtaactatcgtatgcgctgcaatcggcact 363
 DB 279 gtaatacccatatataatcggagctcttggcaactgactagttccctatataatcggtgc 338
 QY 364 ccggaatgagcctcccgagatgaaacaacgtctgtctgctgttcaatgcggaacc 423
 DB 339 ccggaatgagcctcccgagatgaaacaacgtctgtctgtcttaccctctctc 398
 QY 424 gggatgggagctgtctgtctgtctgcacccggaggtgacgttggttctgggagct 483
 DB 399 ctaccctgtctgcacatcgtctatagtgaggccggaagacaaggttgaacagtc--- 455
 QY 484 ggttggtctctgtaaccgcgcgtctgcagcccggaagcttcatgtatgactcgcg 543
 DB 456 -----taccctcccttgagggaggaactatccacccttgagctctcgtaactaac 509
 QY 544 atttcgaggttcaactgtctggtgctctcctgcatactgagcgagatcaatgatacag 603
 DB 510 attctccttaccatagcaggtgtctcctctatctttaggggcataatttcacaca 569
 QY 604 accctctgaacatgacgagcccgagctgacgctgcacaaagtgaggttctctgag 663
 DB 570 acaattatcaataataaacccttcgcatataccaataacgaagccctctctcgtcga 629
 QY 664 tggatcttatcacgagctgtgctgtaacctgtgagcgctgcgctgtctgtgcaatc 723
 DB 630 tccgttcctaatcacagcagctctactctctctatctctccagctcctgctgcagtc 689
 QY 724 accatgtgctgacgagccggaatcttcgacagaccctctcactcctgtctgagcggt 783
 DB 690 actatactactaacagacagccgcaaccctctcacttctgcagcccgaggagga 749
 QY 784 gaccggaattctgtacaaacacatcctgtgcttctgtggacccggaggttacaatc 843
 DB 750 gaccgaattcttacaacaacatctcgatcttcggtacacctggaagttaattctt 809
 QY 844 attctgcccgtcttgacatcacgacatgctgtctgacattctc---gaaaagcg 900
 DB 810 atcttaccaggtctcgatgaataatctcccatcttgtaacttactctcggaaaaaaga 869
 QY 901 gctctcggttacctccgagatgctatgcaatgctggaacatcggtgtctgtgttc 960
 DB 870 ccatttgatgatagtatgtctgagatgataatgataatgttctctgggtttac 929
 QY 961 gcttgagcgcacacatgataacgcttgatgctgcgacccaagcaatctactctg 1020
 DB 930 ggttgagcgcacacatataattacagtaagaaatagacgtagacaagacatattcac 989
 QY 1021 ctggccacatgltgacgagctgctgcgacgagcatlaagcttctctgtgacgcag 1080
 DB 990 tccgtcacatataatcgtctatcccccacgctgcacaaagtatttagtgcctgcaca 1049
 QY 1081 atgttgagcgcctgctgtgtgagttcaaatcgcgagatgctctgtgcttatgttc 1140
 DB 1050 ctcacaggaagaataatgaaatgactgtctgcaagctctgagccctagattcaatctt 1109
 QY 1141 cgttcaacgttggtggtgtgacggtatgctgctgcgacgaagcggtctgagacgtgca 1200
 DB 1110 ctttccacgttggtggtgtgacggtatgctgctgcgacgaagcggtctgagacgtgca 1169
 QY 1201 tatcaagacacatlaacgttggtggtgacacatcaatgagtgtgctgtgtgtgagtc 1260
 DB 1170 ctacaagacagtaactagctgtgagccacttccactatgctcctaataatgagctgta 1229
 QY 1261 ttgtgcatcttccggtatctactttaaagccgaagttctcgggcgagcttcccg 1320
 DB 1230 ttgtgcatatagggagcttacttcaatgatttccctacttccgaagctaaccttgcag 1289
 QY 1321 gaatggcgcgaacgtgacactctgcagcttctcactcgcgtgacgaagctgcttc 1380
 DB 1290 caaactaagccaataatcatttcaactatcaatcatcgcgcgtgtaaatcctaattcttc 1349


```

FT      /*tag= y
FT      /note= "Represented in specification as m, ttt derived
FT      from the entire mitochondrial genome sequence,
FT      residues 6362-6364, EMBL accession no. D38112"
FT      1140..1142
FT      /*tag= z
FT      /number= codon 357
FT      /note= "Position of GTA>GCA, Val>Ala mutation in AD
FT      patients"
FT      1176..1178
FT      /*tag= aa
FT      /number= codon 369
FT      /note= "Position of GAC>GCG, Asp>Gly mutation in AD
FT      patients"
FT      1314..1316
FT      /*tag= ab
FT      /number= codon 415
FT      /note= "Position of ACT>GCT, Thr>Ala; ACT>ATT, Thr>Ile
FT      mutation in AD patients, position of mutation
FT      hotspot"
FT      1317..1319
FT      /*tag= ac
FT      /number= codon 416
FT      /note= "Position of ATC>ACC, Ile>Thr mutation in AD
FT      patients"
FT      1437..1439
FT      /*tag= ad
FT      /number= codon 456
FT      /transl_except= pos: 1437..1439, aa:Met (possible mutant)
FT      /note= "Position of GTA>GCA, Val>Ala; GTA>ATA, Val>Met
FT      mutation in AD patients"
FT      1467..1469
FT      /*tag= ae
FT      /number= codon 466
FT      /transl_except= pos: 1467..1469, aa:Met
FT      /note= "Position of ATA>ACA, Met>Thr mutation in AD
FT      patients"
FT      1473..1475
FT      /*tag= af
FT      /number= codon 468
FT      /transl_except= pos: 1473..1475, aa:Met
FT      /note= "Position of ATA>GTA, Met>Val mutation in AD
FT      patients"
FT      1491..1493
FT      /*tag= ag
FT      /number= codon 474
FT      /note= "Position of GAA>GGA, Glu>Gly mutation in AD
FT      patients"
FT      1581..1583
FT      /*tag= ah
FT      /number= codon 504
FT      /note= "Position of ACA>GCA, Thr>Ala mutation in AD
FT      patients"
XX      WO9526973-A1.
XX      12-OCT-1995.
XX      PD
XX      30-MAR-1995: 95MO-US04063.
XX      PF
XX      03-MAR-1995: 95US-0397808.
XX      PR
XX      30-MAR-1994: 94US-0219842.
XX      PA
XX      (GENE-) APPLIED GENETICS INC.
XX      PI
XX      Davis RE, Herrnstadt C, Miller SM, Parker WD;
XX      DR
XX      WPI; 1995-358577/46.
XX
Query Match      23.7%; Score 397.2; DB 16; Length 1738;
Best Local Similarity 58.1%; Pred. No. 1.6e-96;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

```

```

QY      244 aacggacaccctgtggaacgtcatgttaccacatgttatctcgtatgttctctgtg 303
DB      219 aacggacacatcctaacaagctatcgttcacagcccatgtatctgttaataatcttccata 278
QY      304 ggtatcccccgaatgtctgtgtgtttgtgtaactatctatgacgcgtcgtcaatcgcgct 363
DB      279 gtaataccatcataatcattgagagcttggcaactgactgcttccctataatcgtgtgac 338
QY      364 ccggatattgaccttccgcggtatgaaacaacctgtcgtctgtcgtttcaatgccgtgtacc 423
DB      339 ccggaataggcgttcccccgaataaacaactaagctcttactcttaccctccctctc 398
QY      424 ggaatgggcgtgtgtgtgtgtcttcacacggcggtgacgctgagctgtgtgtggtgtg 483
DB      399 ctactcctgtctgcacatcgcgtatagtgtgagcgcggaagaaagattgaacagctc--- 455
QY      484 ggttggctctgtaccgcgcgtgtcgaccgcgaagctgtctatctcgtatgacctcgcg 543
DB      456 -----taacctcccttagcaggggaactatctccacctgtgagcctcgttagacctaac 509
QY      544 atttcgcggttcaactgtctggtgtggtccctccgatacatgtgcggaatcaacatgatac 603
DB      510 attcttccttaacacctagcaggtgtctccctctatctttaggggcataatcttcatcaca 569
QY      604 accctcttgaacatgcgcgcgcgcgcgcgaatgaacgtgacgaagaatgtgtgtgtgtgtg 663
DB      570 acaatattcaataataaaccctcgtccatcaacacataaccacaagccctcttctgtcga 629
QY      664 tggatcttatacagcgtgtgtgtgtatctcgtgtgcgtgcgtgtgtgtgtgtgtgtgtgt 723
DB      630 tccgtcccaatcacagcagcttctactctctcatctccacagctcgtgtgtgtgtgtgt 689
QY      724 accatgctgtcgtgacgcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 783
DB      690 actataactactcaacagcagcgaacactcaacacacacttcttcgacccgcggaaggaaga 749
QY      784 gaccggaattctgttaacaaacaatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 843
DB      750 gaccocatcttaacaaacaactatctcgtatcttccgttccgtaacccgtgaattatctt 809
QY      844 attctgcgcgcgttctgtgatacatcagccatgtcgtgtgtgtgtgtgtgtgtgtgtgtgt 900
DB      810 atccctaccaggtcttggaataatctcccatattgtaacttactctccggaataaagaa 869
QY      901 gctctcgttctacccgcatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
DB      870 ccaatttgatacatatgatagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 929
QY      961 gcttggtgcgcaacacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
DB      930 ggtgtgagcaccacataattttagtagtaagtagagcagtagacacgagcatattccacc 989
QY      1021 ctggccacacatgtgtatcgcggtgtgcgacgcggaattaaatctctcgtgtgtgtgtgc 1080
DB      990 tccgctacacataatcgtatcgtatcccaacgcggtcgaagaattatgaatgtgtgtgcaca 1049
QY      1081 atgttggtgcggtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
DB      1050 ctccaacggaagcaataatgaaatgatactgtcgaatgtgtgtgtgtgtgtgtgtgtgtgtgt 1109
QY      1141 ctgttaccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
DB      1110 ctttcacacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1169
QY      1201 tatcaagacacataatgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
DB      1170 ctacacagacacgttactatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1229
QY      1261 ttgtgcatcttcgcggtatcttaacttcaatgtccgaagttctcgtgtgtgtgtgtgtgt 1320
DB      1230 ttgtgcatcattagaggtcttcaatcgtatctcccatcttccatcgtacgtacacccctagac 1289

```

Oy	1321	gaatgggctcgcagaagctcgacattccgtgaccttctcatcggctgcgaacgtaactcttc	1380
Db	1290	cnaaacctcaaggccaataaccatttcaactatcatatcatcacgcgttaaattcaactttctc	1349
Oy	1381	ccgcgaagcattcttcgggaacgtcaaggtaatgccgcgcgttatcatcgaactaccgaagcc	1440
Db	1350	ccacaacacttctcgtgcctatccgaaatgccgacgcttaactcgtgatacccgatgca	1409
Oy	1441	ttagcgctgttgagacaagaatctcgcataatggttgtcttcctcgacctcgcctctgtctctg	1500
Db	1410	taccaccaagaagaacatccatcatcatctgtagctcatcattctctlaacagcaqtaata	1469
Oy	1501	ttctcatcgtgatcttt	1518
Db	1470	ttaatatttctcatgatt	1487
RESULT	5		
AAZ11492	ID	AAZ11492 standard; DNA; 16569 BP.	
XX	AC	AAZ11492;	
XX	DT	02-NOV-1999 (first entry)	
XX	DE	Mutated human mitochondrial DNA sequence.	
XX	KM	Mutation: human: mitochondrial gene; abnormality detection: diagnosis;	
XX	OS	diabetes mellitus; ds.	
XX	OS	Homo sapiens.	
XX	OS	Synthetic.	
FM	Key	Location/Qualifiers	
FT	mutation	3423	
FT		/+tag= a	
FT	mutation	/note= "native G is mutated to T"	
FT		3426	
FT		/+tag= b	
FT		/note= "native A is mutated to G"	
XX	PN	JF11221077-A.	
XX	PD	17-AUG-1999.	
XX	PE	04-FEB-1998; 98JP-0023196.	
XX	PR	04-FEB-1998; 98JP-0023196.	
XX	PA	(SAKA) OTSUKA PHARM CO LTD.	
XX	DR	WPI; 1999-520716/44.	
PT	PT	Mutated human mitochondrial gene - useful for diagnosing diabetes	
XX	PT	mellitus	
PS	PS	Disclosure; Page 14-19; 27pp; Japanese.	
CC	CC	The invention relates to a DNA containing at least Base No. 3426 site of	
CC	CC	a mutated human mitochondrial gene. The DNA contains a mutation at base	
CC	CC	3426 where adenine is replaced by guanine or the base is deleted. The DNA	
CC	CC	maintains identification specificity of human mitochondrial gene. The	
CC	CC	invention provides a method for the detection of abnormality in human	
CC	CC	mitochondrial gene in which the replacement of the base from A to G at	
CC	CC	the site of Base No. 3426 in the human mitochondrial gene of a patient is	
CC	CC	detected. A reagent kit containing a primer that can be used for the	
CC	CC	synthesis of the mutated human mitochondrial DNA can be used for the	
CC	CC	diagnosis of diabetes mellitus. The present sequence represents a	
XX	XX	mutated human mitochondrial DNA.	
XX	XX	Sequence 16569 BP; 5121 A; 5175 C; 2178 G; 4095 T; 0 other:	

Query Match	23.7%;	Score 397.2;	DB 20;	Length 16569;
Best Local Similarity	58.1%;	Pred. No. 3.9e-96;		
Matches 743;	Conservative 0;	Mismatches 523;	Indels 12;	Gaps 2.

QY	244	aagcgacacgttggaagcctcagcttcttaccacacatggtatctcgtatgcatgattctcttg	303
Db	6051	aacggaacacatctacaagcgttatctcgtacacagcccatgcatcttgtaataatctcttcata	6110
QY	304	ggtatcccgcatctgtctcggttggttttgtaatactctgatatctgatgcgctgcaaatgcgcgt	363
Db	6111	gtaataccacataatactcgagagctttgttgcaactgactagttcccttaantaatcgtgcgc	6170
QY	364	ccggaatagtccttcccggtatgaaacaacctgctgtcttcgttcatttgccggaac	423
Db	6171	cccgataatgctgttcccgcaataacaacaataagctctctgactcttaacctccctctc	6230
QY	424	gcgattgagcgctggtctcgtctggttctgcacacggcggtgacgctgcagctggtttcggcgct	483
Db	6231	ctactcgtcgtcactctgtatagtgagagcccggaagcaagaacggtttgaaagct---	6287
QY	484	ggttggtgtctgtacccgcgcgtgtgcacccggaagctgtgatttcgatagagccctgcg	543
Db	6288	-----taccctcccttaagaggaactactctccaccccttgagacctccgttaactaac	6341
QY	544	atttgcggttactgtgctgggttgctctctgcatcagggcgagatacaatgatacag	603
Db	6342	atcttcccttaaaccttaagaagtgctctctctatcttaaggccataattatcata	6401
QY	604	acctcttgacacatgctgcgcccccggaacagacgtctgcacaagaatgctgtttctctg	663
Db	6402	acaatlaataataaaccctcgcatcaaccacaaataccaagaagccctcttcgtctga	6461
QY	664	tcatcttatacagcgttggtgctgataccctcgtcgcgctgcggttcttgctggtgcacatc	723
Db	6462	tcgcctcctaatacgaagtcctctaactctccctacatctctccagctctagatgcgtgcac	6521
QY	724	acgatgtgttgacagacccglaactctgcgcagaccttcttaactcgtctgcgcgcgt	783
Db	6522	actatactactaaagacccgcaacctcaacacacacctcttcgtacccgcggaggagga	6581
QY	784	gaeccgatctctgacaacaacatcctgvtgtctctttggcacccggaagtatacatc	843
Db	6582	gaecccatctatacaacaacccctatctgtatcttctgcacacccgaaagtataatctct	6641
QY	844	attctgcgcgcttggtgacatcagcagctctgctgctgacacctctc---gaaaaagccg	900
Db	6642	atctcaacaggtctcggaataatctcccatattgtaacttactactccggaaaaaaagaa	6701
QY	901	gtcttcggttacctgcgcgaatggtctatgtaaatgtgtaacatcgtgtgtctggtcttgctc	960
Db	6702	ccatttgataatagttgattgctgtcgtatgataatataatgtgcttccctaggtttatc	6761
QY	961	gtctggtgcgacacacatgatacacccgttggtatgctgcgtgacccagcaatctactaatg	1020
Db	6762	gtgtgagcaacacatataatttaagataggaataagacgtatagaacaacgagcatattccac	6821
QY	1021	cttgcacacatggtatgctgcgtgtgcgcagccggaatgaatatctctgtgatacgcacg	1080
Db	6822	tcgctacataataatcatcgtctatcccaacccgcggtctaaagaattagctgactgcgcaca	6881
QY	1081	atgtgtggtgcgctcgtgttgattgattcaaatcgcgcgaatgctcttggtgcttggcttatgctc	1140
Db	6882	ctccacgaagcaaatatgaaatgattctgctgtgcagctgctcgtgcacctagattatcattt	6941
QY	1141	ctgttaacccgttggttggttgtagaccggtatccgtgtgccccaaagcggtctgagccgtgtca	1200
Db	6942	cttttccacgttaggtgtgcgcgtgactggtcatgttatgagcaaacctcactagaacctgtta	7001
QY	1201	tatcaacacacttaattcgtgtgcgcgcacattccatcatgtatgttcgctgtggttgagctc	1260
Db	7002	ctacaacgaacgatactacgtttgtagcccaacttccactatgctactcaatatagagcgttta	7061
QY	1261	tttgcgatacttcgcgcggtatcatcatttatagcgcgaagtctcgtggccgcgcttcccg	1320

Db	7002	ctacacgacacgactacgctgttagtcgccactccactcatgtgctcctacatgagagctga	706
Qy	1261	tttgcgactcttcgcgcgttactactctttaaagcgaattctctgggcgcgcgcgtttccg	1320
Db	7062	tttgccatcaatgagaggtcttcattcactcgtattccctcatctccaggctaacccctagac	7121
Qy	1321	gaatggctgcgaagaagcgacactctctgacctcttcctcatcgtgcgcaacgctcagcttc	1380
Db	7122	caaacctacgacgaataatccattcattcatcatcatcattcatcggtgaattctaactcttc	7181
Qy	1381	ccgcgaacactctctcgggaagctcaagggtatggcgcgcgcgttaactgactatcccgaaac	1440
Db	7182	cccaaaacactctctcgcgcctaccgcgaatggcccgacgcttactcgtgacatacccgatga	7241
Qy	1441	ttcgcgcgtctggaagaagatcctgctccatagtgtgcttctctgacctcgcctcgttctcg	1500
Db	7242	tacacccatitgaacaatcatcatcatctgtagtgctcatctatcttctcctaaacagcagtaata	7301
Qy	1501	ttcttcacatcgatgacttt	1518
Db	7302	tttaataatttcacatgatt	7319

RESULT 8

ID AAZ57203 standard; DNA; 16569 BP.

AC AAZ57203;

DT 30-MAR-2000 (first entry)

DE	Human mitochondrial DNA séquence SEQ ID NO:2
DE	Human mitochondrial DNA séquence SEQ ID NO:2

KM Human, mitochondrial DNA; extramitochondrial DNA; mtDNA; exmtDNA;
 KM diagnosis; quantification; detection; dystonia; Alzheimer's disease;
 KM Huntington's disease; Parkinson's disease; schizophrenia; stroke;
 KM non-insulin dependent diabetes mellitus; mitochondrial encephalopathy
 KM lactic acidosis; myoclonic epilepsy ragged red fibre syndrome;
 KM Leber's hereditary optic neuropathy; ds.

OS Homo sapiens.

PN W09966075-A2

PD 23-DEC-1999.

PF 14-JUN-1999; 99WO-US13426.

PR 15-JUN-1998; 98US-0097889

PR 30-APR-1999; 99US-0302681.

PA (MITO-) MITOKOR.

PI Herrnstadt C, Ghosh SS, Cleveneger W, Fahy ED, Davis RE,

DR WPI; 2000-097754/08.

PT Quantification of extramitochondrial DNA for diagnosis

PS Claim 29; Fig 2; 157pp; English.

CC The present invention describes a method for the quantification of
CC extramitochondrial DNA (exmtDNA) by determining the ratio of a first
CC and second biological sample containing exmtDNA and mitochondrial DNA
CC (mtDNA) to determine the risk or presence of a disease associated with
CC altered mitochondrial function. The method can be used to determine
CC the risk of or presence of a disease associated with altered
CC mitochondrial function, especially Alzheimer's disease, Huntington's
CC disease, Parkinson's disease, dystonia, schizophrenia, non-insulin
CC dependent diabetes mellitus, mitochondrial encephalopathy, lactic
CC acidosis, stroke, myoclonic epilepsy ragged red fibre syndrome and

CC Leber's hereditary optic neuropathy. The method can also be used to
CC identify agents suitable for treating such diseases, in particular
CC Alzheimer's disease. AAZ57202 to AAZ57313 represent nucleotide sequences
CC used in the exemplification of the present invention. More specifically
CC AAZ57206 to AAZ57313 are PCR primers used in the detection of exmtDNA
CC and mtDNA.
XX
XX
SQ Sequence 16569 BP; 5122 A; 5180 C; 2171 G; 4096 T; 0 other;

Query Match	23.7%;	Score 397.2;	DB 21;	Length 16569;
Best Local Similarity	58.1%;	Pred. No. 3.9e-96;		
Matches 743;	Conservative	0;	Mismatches 523;	Indels 12; Gaps 2.

QY	244	aagcgaacccgtcgtgagaaacgtacgtatctacccacagatctatctatgattgtctctgtg	303
Db	6051	aagcgaacacatctaaacacgttatctgtccacagcccaagatcttgataatactcttcata	6110
QY	304	ggtatcccgcatctgttcggatgttttgtaatactatctgatactgcgcgtcyaataatcgcgct	363
Db	6111	gtaataaccctcatalaactcggaaagccttttgcgaactgaactgtctccctataatcgtgtcc	6170
QY	364	ccgatatctgccttcctccggctatgaaacacacgtctgtctgtctgtctatctgcggtaac	423
Db	6171	ccgatatctgccttcctcccgcaataaacaataaagctctgtactcttaaccctctctc	6230
QY	424	gcgaatggcgtgagcttcgctgttcgcacccggcggtgacgctacgctggtgttcggcgct	483
Db	6231	ctactcctgcctgcgatctgtcatatgtagggccggagccgagaaacagtttgaacagctc---	6287
QY	484	ggttgaggtctctgtaaccggcgctgtcgaccgcggaaagcgtctatctgatactgactccg	543
Db	6288	-----taacctcccttagcaggaaactatctcccaacctgcggacctccgtatgacctaac	6311
QY	544	attctcgcggtctcaactgtctcggtgtgcactctcgatctatgtggcgtatacaatgatcacg	603
Db	6342	atctctctcttaacacactagaggtgtctctctatctcttaagggccataatcttcatac	6401
QY	604	aactcttgaaacatgctgcgcgcccgacatgaaagctgacaaagttgcgtgttctctgtg	663
Db	6402	acaattatcatataaanaacccctctcatataaccaataacaaagccctctctgcgtga	6461
QY	664	tgcattcttaacagagctcttgctgcgtacccgtcgtgcgtgcgtgtctgtcgtgtaacatc	723
Db	6462	tcgctcttaacaaagagctcaactctctctctatctctctccagctcctagctcgtgtgaac	6521
QY	724	accatgtcgtctgacagcgcgtaactctgcgaacgacctcttcatactcgtctgcgcggt	783
Db	6522	actatactactaaacagccgcaacactcaacacacactcttctgcaccccgcgagagga	6581
QY	784	gaccgcatctgtacccaacacatccgtgtgtctctttgtggcacccggaggtgaacatc	843
Db	6582	gaccacatctataacacaacacctatctcgtattcttcgcgcacccggaagttatactctc	6611
QY	844	attctgcgcgctttgcatcatcatcaagcatalctgtgtcgaacctctc---gaaaaagcgg	900
Db	6642	atctcaacaggcttcggaaataatctcccatattgttaacttactactcggaaaaaagaa	6701
QY	901	gtctcttgatctacccgcggaatggtctatgtcaatgtgtgcaatcgtgtctcgtgcctttgc	960
Db	6702	ccatttgatgatacagaatgatagtgtctgagctatgataatcaatctggtctcctaggtttac	6761
QY	961	gtctcgtggcacccacgtgataacacggtttgattgtcgtgacccagcaatctcaattatg	1020
Db	6762	gtgtgtagaacccctataattattacagttagaataagacgtgaaacacgagatatttcc	6821
QY	1021	ctggacccaatgtgatactcgtgtgcgcgacccgatatgaatctctctcgtgtgatacgcag	1080
Db	6822	tcgctatacatatcatcgtcatctatcccaacccggtctaaagtatttagctgtactgcaca	6881
QY	1081	atgtgtgggcggtcgtgtgtgattgattcaaatcgcgaatgctctcgtgcctttggtcttatgtc	1140
Db	6882	ctccacgggaagcaataagaatgatctcgtcgcagtgctctgttcggcccttaggtatctcttc	6941

DB 695 TCCTGATTTCCGTTCTTACCGCGGCTTAAGTCTTCAACCTGGGTTGGAAACATGTGAAG 636
 QY 1091 gctcggtagtgaagcaaatgcgcgatgctctggsccttggcttcttctcgttccacg 1150
 DB 635 GTCACTACCTCTGGGAAACCCCAATGATGTGCTGTGGCTTCAATGGCTACCTCTCT 576
 QY 1151 tgggtgtgtgacccgtagcgtctgcccgaagcggtctgacgctgcatatcacaca 1210
 DB 575 TCGGTGTGTGACCGGCAATATATGTGGCTGCCCAACACTGAGACTTCCACTTGGCTACT 516
 QY 1211 ccattatcgttgtagcgaactccattatgtagtctgctggtagcgttcttgcgact 1270
 DB 515 CCTACTTCTGTGATCGCGCACTTCCACTACACCTCTTCCGTTACCGTGGTGTGCGATCGT 456
 QY 1271 tggcggatcttaactttacatgacgaagttctggcgccgcttccggataggcgt 1330
 DB 455 GTTCAGGCGCTTACTCTTCTGTTCCGAATGATGACTGGCCGATGATGACGAGCGCTCTTG 396
 QY 1331 caaagctgacactctgacactcttccatcgtgtagcgaactgacttctccgcgacact 1390
 DB 395 GCAAGATTCACCTTCTGTGACTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 336
 QY 1391 tcttggaagctgaaggtatgctgctgctgcttcaatcgtact-----atccgaagccttcg 1444
 DB 335 GGGTGGGCAACATGGGTATGCGACGCTTACGCTGACCTGATGCTGATGCTGATGCTGAT 276
 QY 1445 cccttggaagcaagctgctgcttctgctgctgctgctgctgctgctgctgctgctgct 1504
 DB 275 CCATTTACACCAAGATTCACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 216
 QY 1505 tcaatcgtatcttctgcttctgctgctgctgctgctgctgctgctgctgctgctgct 1564
 DB 215 TCATCTGGAACGCTTCAAGTCTCTGCGCTACGGTGAGCTGATGCTGATGCTGATGCTGAT 156
 QY 1565 gggggaatctgcgcgaatgctggaatggaatggaatggaatggaatggaatggaatgga 1624
 DB 155 GGGG---TTACGGCACTCTCTGAGTGGGCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 99
 QY 1625 tcgaagcgtcgc 1637
 DB 98 TCGCATCTTGGC 86
 RESULT 12
 AAH67734
 ID AAH67734 standard; DNA; 1743 BP.
 AC AAH67734:
 DT 26-SEP-2001 (first entry)
 XX C glutamicum coding sequence fragment SEQ ID NO: 2769.
 DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 KW Corynebacterium glutamicum.
 OS
 XX
 PN EP1108790-A2.
 PD 20-JUN-2001.
 PF 18-DEC-2000; 2000EP-0127688.
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 PA (KYOW) KYOMA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.
 DR P-PSDB: AAG92515.
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 -
 Claim 8: SEQ ID NO: 2769; 246pp + Sequence listing; English.
 The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 XX
 SQ Sequence 1743 BP; 283 A; 541 C; 427 G; 492 T; 0 other;
 Query Match 17.7%; Score 296.6; DB 22; Length 1743;
 Best Local Similarity 54.0%; Pred. No. 1.4e-69;
 Matches 752; Conservative 0; Mismatches 614; Indels 27; Gaps 6;
 QY 251 accttggaagctatggttactacatggtatcttgatgattcttctggtatcc 310
 DB 224 agcagttcaacagctgttccacatgacgaagctgcatgctgtctgtacgaa--- 280
 QY 311 ccgcatgttctggtgttttggtaactatctgtagcgcgcgcgaatgagcgcccgata 370
 DB 281 ctccaattgtttgggttttggtaactatctgtagcgcgcgcgaatgagcgcccgata 340
 QY 371 tggcctccgcgtagaacaacactgctgctgctgctgctgctgctgctgctgctgctgct 430
 DB 341 tagcttccacagctttagatgcttctgctgctgctgctgctgctgctgctgctgctgct 400
 QY 431 gcttgcttgcgttctgacacggcggtgacggtcagctggtgctggtggtggtggt 490
 DB 401 tctgacccgcttctgacacggcggtgacggtcagctggtgctggtggtggtggtggt 460
 QY 491 tctgtacccgcttctgacacggcggtgacggtcagctggtgctggtggtggtggtggt 550
 DB 461 cactgtcgaacgaattcaatccccaaggccttgcctt-----gacatgtgattgtcgt 514
 QY 551 cgttcaactgtcgggtgctcctcctgacatgtagggcgatcaacatgatacgaacttct 610
 DB 515 gttctggtgacacgtgatttgcctcgttgccttgcgaatataacatgatacgaacttct 574
 QY 611 tgaacatgacgcccccggaatgacgctgacaaagtgcgtgttctctgtgtgactc 670
 DB 575 tctgcttcgcgcaactgtgatacgaatgctcgtatgacatttcaactgtgatact 634
 QY 671 ttatacggcttgcgtgactcctgctgctgctgctgctgctgctgctgctgctgctgct 730
 DB 635 tctgttctgcttctgtctgctgctgctgctgctgctgctgctgctgctgctgctgct 694
 QY 731 tctgacccgctgaactcgcgacgaacttctcaactcctgctgctgctgctgctgctgctgct 790
 DB 695 tctgtatgacccgcaagcttggtagaacctgtatgatacgaactcgaagggcgctccc--- 751
 QY 791 tctgtaccaacacatctcgttcttcttgggaaccccggaagtgtatcaatcaatctcgc 850
 DB 752 tctgtgacgaacactgtctgcttctcgtgacacccctggaaggttctgcttgcgctgc 811
 QY 851 ccgcttgcgacatcatcagccatgtctgtgcgaacttctcgaaaaaacgcttcttgcgt 910

Db 812 cgtcttcggcattgtcttcgagatcattccgtgtcttcctccgtaagcaaatgttcgtt 871
 QY 911 acctgcgaatggtctatgcaatggtggcaatggtgtcttcggccttgcgttcggcgc 970
 Db 872 agctgcgcgtatcttcgcaactgttcacatgtgtgcaactgacatgagctgtgagctc 931
 QY 971 accacatgtaacacgttgtgtatgtcgttcacccagcaatcctactcatcgtctgacaca 1030
 Db 932 accaatgttctgaactgtggcgagtttgtcttcctcgtcttcctcctcag-----agct 985
 QY 1031 tggtagtcggtgtgcgcgaacgcatgaatctctcgtgtgatalccgacgatgtgggctc 1090
 Db 986 tctgtatcttcgtcttcacgcggttaagtcttcacacgtgtgtgaaacacatgtgaaag 1045
 QY 1091 gctcgttgaatgtaacatgcgcgcgttcgttcggttcgttcgttcgttcgttcgttcacg 1150
 Db 1046 gtacactcaactgtggaaaccccaatgcatcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1105
 QY 1151 tgggtgtgtgtgacgcgtatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1210
 Db 1106 tctgt 1165
 QY 1211 cctatagctgt 1270
 Db 1166 cctactcctgt 1225
 QY 1271 tgcgcgtatcacttcttcacatgcaatgtctgcgcgcgttcgttcgcgcgttcgcgcgttc 1330
 Db 1226 gtgcagcgcttactctgt 1285
 QY 1331 caaagctgactctgtgacactcttcacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1390
 Db 1286 gcaagatcactctgt 1345
 QY 1391 tctgt 1444
 Db 1346 ggt 1405
 QY 1445 cgt 1504
 Db 1406 ccatcacaacagatcttcacacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1465
 QY 1505 tcatgt 1564
 Db 1466 tcatgt 1525
 QY 1565 ggggcgaatcgcgcgaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 1624
 Db 1526 gggg---ttacggcaactcctcgtgagtgggcaactcctcgtcctcctcctcgcacaact 1582
 QY 1625 tcgaaacgctgc 1637
 Db 1583 tcgcatcctgtgc 1595

RESULT 13
 AAH68532/c
 ID AAH68532 standard: DNA; 349980 BP.

AC AAH68532;
 XX 26-SEP-2001 (first entry)
 DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
 XX
 KW Corynebacterium: amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.

PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KJOW) KYOMA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI: 2001-376931/40.
 XX
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT

Disclosure: SEQ ID NO: 7067; 246pp + Sequence Listing: English.
 CC
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from corynebacterium bacterium, and identifying a homologue of a gene derived
 CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

Sequence 349980 BP: 80900 A; 98397 C; 92139 G; 78544 T; 0 other;

Query Match 17.7%; Score 296.6; DB 22; Length 349980;
 Best Local Similarity 54.08; Pred. No. 1e-68;
 Matches 752; Conservative 0; Mismatches 614; Indels 27; Gaps 6;

QY 251 acctgtgaaagcatatggttacatcagatgattcctgatatgtcttctgtgtgtatcc 310
 Db 272582 AGCAATTAAACACACTGTTCACATGACGAGAACTGTATCTCTCTGTACGGA--- 272524
 QY 311 ccgaatgttctgt 370
 Db 272525 CTCGAATTGTTGGGGTTCCTAACAACGTCGCACTTCAAGATGGTGCCCTGACG 272466
 QY 371 tggcttcgccggtatgaaacactgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 430
 Db 272465 TAGCTTTCACAGTTCATTAACCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCT 272406
 QY 431 gctgt 490
 Db 272405 TGCTGACCGGCTTCCTGACCCCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 272346
 QY 491 ttcgtacccgcgt 550
 Db 272345 CACGTGTGACGCAATTCACCTCCAGCCCTTGCTCT-----GACATGTGATTTGTG 272292
 QY 551 cgtgtacactgt 610
 Db 272291 GTGTGCTGCAACTGTGATTTGGCTTCGCTTCCTTCGCAATTAACATCTTCACCACTCC 272232
 QY 611 tgaacatgctgc 670
 Db 272231 TCTGCTCTCGCGCACCTGTGATGACATGTTTCGTTATCTTATTTTCACCTGGAATATCT 272172
 QY 671 ttatcacggtctgt 730

CC acid, alcohol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).

Sequence 1779 BP; 294 A; 548 C; 435 G; 502 T; 0 other;

Query Match	17.68;	Score 295;	DB 22;	length 1779;
Best Local Similarity	53.98;	Pred. No. 3.8e-69;		
Matches 751; Conservative	0;	Mismatches 615;	Indels 27;	Gaps 6

[illegible][illegible]

```

RESULT 15
AAH84533
ID AAH84533 standard; DNA; 1992 BP.

```

AC AAH84533;

DT 26-SEP-2001 (first entry)
 YY

DE E. coli growth and proliferation related coding sequence SEQ ID NO:161.
XX

Escherichia coli; growth; proliferation; microbial; antimicrobial;
(bacterial) infection; microorganism; de

Escherichia coli.

PN W0200134810-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30950.

PR 09-NOV-1999; 99US-0164415.

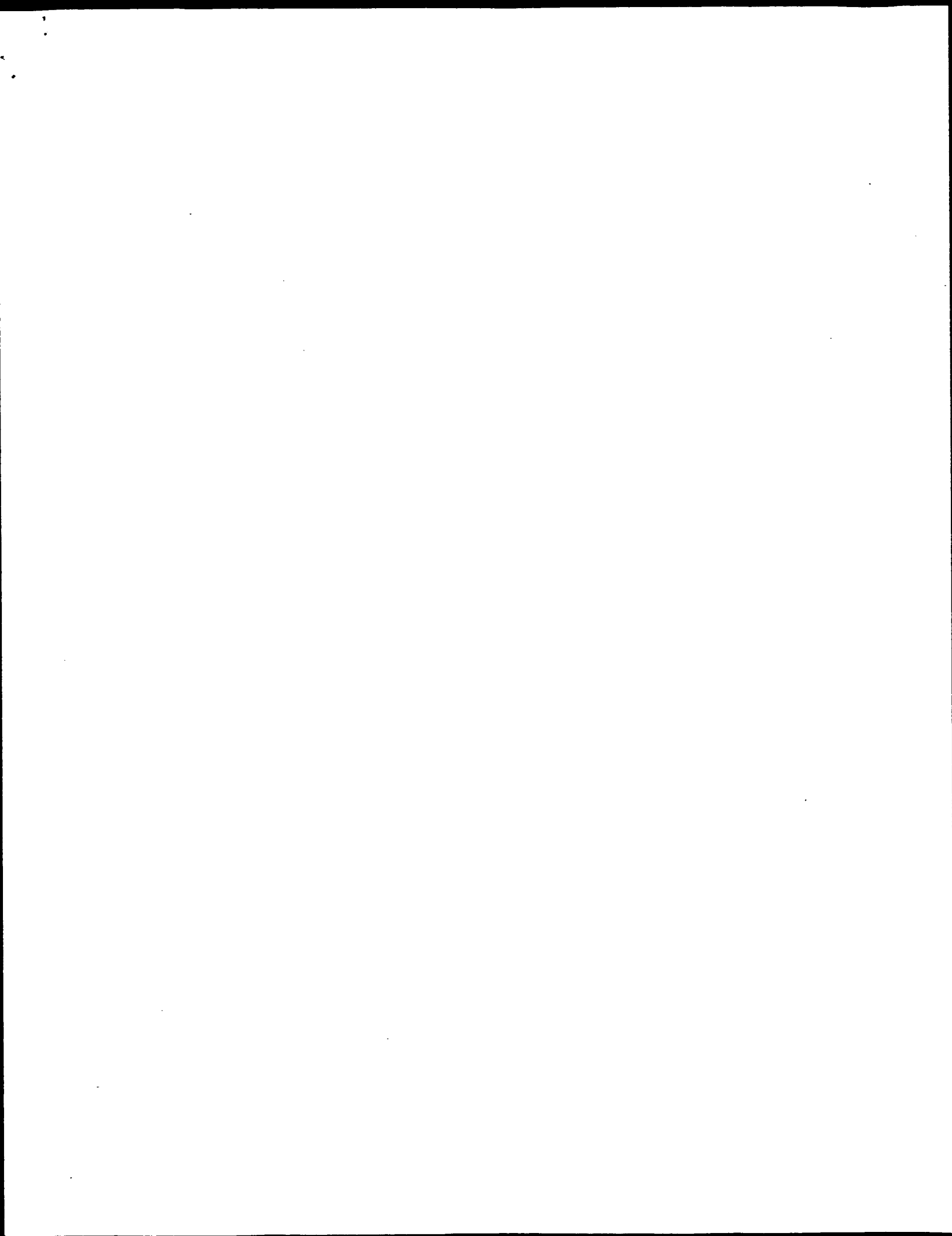
PA (ELIT-) ELITRA PHARM INC.
XX

Forsyth RA, Ohlsen K, Zyskind J,
PI
xy

DR WPT; 2001-335933/35.
DR P-PSDB: AAG9886?

Novel nucleic acid

Novel nucleic acids that inhibit *Escherichia coli* proliferation, useful



GenCore version 4.5
Copyright (c) 1993 - 2000. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 22, 2001, 08:09:49 ; Search time 113.93 Seconds

(without alignments)
3327.694 Million cell updates/sec

Title: US-09-712-768-1

Perfect score: 1674

Sequence: 1 atgacgacgcgcgcacatca.....acaagcctccctgcacata 1674

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2-6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2-6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2-6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2-6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2-6/ptodata/2/ina/PCUTS_COMB.seq:*
6: /cgn2-6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397.2	23.7	1613	1	US-08-219-842-1
2	397.2	23.7	1613	1	US-08-451-096-1
3	397.2	23.7	1613	2	US-08-810-599-1
4	397.2	23.7	16569	4	US-09-097-889-2
5	388.2	23.2	6744	4	US-09-097-889-1
6	384.4	23.0	1735	4	US-08-413-740A-1
7	384.4	23.0	1735	5	PCT-US95-04063-1
8	147.2	8.8	773	4	US-08-998-416-564
9	141	8.4	484	4	US-09-385-982-343
10	122	7.3	626	4	US-09-385-982-355
11	111	6.6	663	4	US-08-998-416-187
12	111	6.6	719	4	US-08-998-416-1138
13	111	6.6	856	4	US-08-998-416-289
14	108	6.5	277	1	US-08-518-878B-19
15	108	6.5	277	1	US-08-294-522B-19
16	108	6.5	277	2	US-08-807-861A-19
17	108	6.5	277	2	US-08-470-868A-19
18	108	6.5	277	3	US-09-210-681-19
19	108	6.5	277	3	US-08-946-719A-19
20	106.4	6.4	277	1	US-08-518-878B-18
21	106.4	6.4	277	1	US-08-294-522B-18
22	106.4	6.4	277	2	US-08-807-861A-18
23	106.4	6.4	277	2	US-08-470-868A-18
24	106.4	6.4	277	3	US-09-210-681-18
25	106.4	6.4	277	3	US-08-946-719A-18
26	96.4	5.8	632	4	US-09-385-982-507
27	83.8	5.0	4539	1	US-08-119-512-1

28	83.8	5.0	4539	1	US-08-488-015B-1	Sequence 1, Appl
29	82.8	4.9	4542	3	US-08-814-412-11	Sequence 11, Appl
30	75.4	4.2	700	4	US-09-328-111-694	Sequence 694, App
31	69.6	4.5	1819	2	US-08-371-377-20	Sequence 20, Appl
32	69	4.1	577	4	US-09-385-982-203	Sequence 203, App
33	67	4.0	1926	2	US-08-978-182-2	Sequence 2, Appl
34	67	4.0	1926	2	US-09-205-681-2	Sequence 2, Appl
35	66.8	4.0	119	3	US-08-582-261A-6	Sequence 6, Appl
36	66.8	4.0	119	3	US-09-016-540-6	Sequence 6, Appl
37	61.8	3.7	620	4	US-09-385-982-208	Sequence 208, App
38	56.8	3.4	180	2	US-08-526-964-16	Sequence 16, Appl
39	56.8	3.4	180	2	US-08-946-617-16	Sequence 16, Appl
40	50	3.0	50937	4	US-09-428-517-1	Sequence 1, Appl
41	49.4	3.0	7218	1	US-08-232-463-14	Sequence 14, Appl
42	46	2.7	696	4	US-08-998-416-779	Sequence 779, App
43	45.6	2.7	119	1	US-08-582-261A-5	Sequence 5, Appl
44	45.6	2.7	119	3	US-09-016-540-5	Sequence 5, Appl
45	40.6	2.4	612	4	US-09-385-982-280	Sequence 280, App

ALIGNMENTS

RESULT 1
US-08-219-842-1
: Sequence 1, Application US/08219842
: Patent No. 5565323
: GENERAL INFORMATION:
: APPLICANT: Parker, W. D.
: APPLICANT: Herrstadt, Corina
: TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
: TITLE OF INVENTION: for Alzheimer's Disease
: NUMBER OF SEQUENCES: 95
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/219, 842
: FILING DATE: 30-MAR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31, 815
: REFERENCE/DOCKET NUMBER: P-AG 9504
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-8949
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1613 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-08-219-842-1

Query Match 23.7%; Score 397.2; DB 1; Length 1613;
Best Local Similarity 58.1%; Pred. No. 7.2e-99;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

Qy 244 aacgacacccgttgacacgcacgttactcctacacgtatctcgtatgacgtcttgg 303
Db 219 AACGACCATCTACACGCTTATCTGTCACAGCCCATTTGTATATATCTTTCATA 278

OY	304	ggtcccgcatctgttcggttggtttgttgtaactctatcgatgacgcgtgcaactcgagcgt	363
Db	279	GTAAATACCCATCAATAATCGAGGCTTTTGCACTACTACTAGTTCCCTATATATGSGTCC	338
OY	364	ccgagatctgacgtcccgcgatgaaacaacgcgtctgtctgtgttcaattgccgtaac	423
Db	339	CCCGATATGGCGTTTCCCGCATMAACACATTAAGCTTCTGACTTACCTCCCTCTC	398
OY	424	gcgctgagcgttgctctcgctgtgttcgcaacggcggttgagcgtcaagcttbggttcggcgtt	483
Db	399	CTACTCTGCTGCATCTGCTATATGGAGGCGGAGAGCAAGTTGAACAGTC---	455
OY	484	gattggtctctgtaacccgcgcgtctgaccccgagacgtggtctatctgatatgacctcg	543
Db	456	-----TACCCTCCCTTAGAGGAGGAACTACTGCCAACCCGTGGAGCCCTCGTAGACTTAAC	509
OY	544	atttcgcggttcaactgttcgggtgtccctctcgatcaataggcgcgatcaacaatgtaacg	603
Db	510	ATCTTCTCTTACACCGTAGCGGTGCTCTCTATCTTAGGGGCGCATCAATTTATCTACA	569
OY	604	acctctctgacaatgcgcccgcccgcgatcgagcgtcgacaagatgcgtttctgcgtg	663
Db	570	ACATTTATCAATTAATAAACCCCTCGCATTAACCAATAACCAAGCCCTCTTGCTGTGA	629
OY	664	tgcgacttctatcacgcttgcgtatctcgtcgtcgcgctgcggtctctgtctgtgtgcaatc	723
Db	630	TGCCCTCTATACAGAGACGCTACTTTCCTCATCTCTCCCACTGCTAGCTGTGCATC	689
OY	724	accgtcgtctgacacgcgttaactctggagcaagaccttcaatctcgtcgcgcggt	783
Db	650	ACTATACTACTACAGACGCCCAACCTCAACACACGCTTCTTGAGACCGCGGAGAGGA	749
OY	784	gaccgcgactctgtlaacaacaacatcctgttgctcttcttgagcaccggaagtgtacatc	843
Db	750	GACCCCATCTTATACCAACACCTATTTCTGATTTTTCGGTACCCGTGAAGTTTAATCTTT	809
OY	844	atttcgcgagcttgcgacatcatcagcgatgcgttgctgtgtaacctctc---gaaaagcgcg	900
Db	810	ATTCCTACCGGCTTCGGAATAATCTCCCATATTGTACTTACTACTCCGGAAAAAGAA	869
OY	901	gtcttcggttctcccgacatggtctctatgtcaatggttgacatcgtgtctctgtgctttgc	960
Db	870	CCATTGTGATATCAATGAGTATGCTGAGCTATGATATCAATTTGAGTTCTTAGGGTTTATC	929
OY	961	gtctgcggcgacacatgtaacaccttggtatgtgcgtgaaccagaaatcctactatgt	1020
Db	930	GTGTAGACCACTAATATTTTACAGTAGTAGAATAGCTAGACACAGACATATTTTACC	989
OY	1021	ctgagccacaatggtgtaatcgcggtgcgcgacccggatcaaatctctctgtgatacgcaag	1080
Db	990	TCCGCTACCAATATCATTCGGTATTCGCCACGCGGTCAAAAGTATTAGCTGACTCCGACA	1049
OY	1081	atgttgaggcgctcggttgatgttcaaatatcgccgaatgcgtcttggtgccttggctttatgtc	1140
Db	1050	CTCCACGGAAACCAATATGAATATGATCTGCTGCAGTCTCTGAGCCCTTAGAGATTCAATCTT	1109
OY	1141	ctgtctacacggtggtgtgtgtaaccggtatcgtctgtgcgaagcg9ggtatgtgaacggtca	1200
Db	1110	CTTTTACACCGTAGGTGCGCTACCTGGCATTTGTATGACAACTCATCATATACACATCGTA	1169
OY	1201	tatcagcaacttatatcagtgttggtgcgacttccattatgtatgtcgtctgggtgtgatac	1260
Db	1170	CTACACGACAGTACTACTGTTGTAGGCCCATCTTCCATATGTCTCTATACATATAGACCTGTA	1229
OY	1261	tttgagactctgcgcggtatctactctttaaagtcgaagtcttcgtgcgcgcgcttcccg	1320
Db	1230	TTTTGCATCAATAGAGAGGCTTCAATCACTGATTTTCCCTATTTCTCAAGGCTACACCTTAGAC	1288
OY	1321	gaatgagctgtaaaagcttgcaactcttggaacttcttcaatcogtgtgcgaagcttcaagcttctc	1380
Db	1290	CAAACTACGCGCAAAATTCATTTTCACTATATTCATGTCATGCGGCTAAATCTAATCTTCTTTC	1349
OY	1381	ccgagcaactctctggagcgtlcaaggtatgacgcgcgcgcttacaatcgatactccgaagcc	1440

Db	1350	CCACACACATTTCTCGGGCTTACCGGAATGCCCGGACGTACTCGGAGTACCCCCGATGCA	1409
Oy	1441	ttcgcgctgtggaacaaagctctgctctatgtaggtcttcctgacctgcgcctgctctg	1500
Db	1410	TACACCCACATGAAGAACATCATATCTGTAGGCTCATTTATTTCTCTAACAGCAGTAATA	1469
Oy	1501	ttctcaatcgtagatcctt	1518
Db	1470	TTAATTAATTTTCATGATT	1487

```

1      RESULT      2
2      US-08-451-096-1
3      : Sequence 1, Application US/08451096
4      : Patent No. 5760205
5      : GENERAL INFORMATION:
6      : APPLICANT: Parter, W. D.
7      : APPLICANT: Hernstadt, Corinna
8      : TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
9      : TITLE OF INVENTION: for Alzheimer's Disease
10     : NUMBER OF SEQUENCES: 95
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Campbell and Flores
13     : STREET: 4370 La Jolla Village Drive, Suite 700
14     : CITY: San Diego
15     : STATE: California
16     : COUNTRY: USA
17     : ZIP: 92122
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: PatentIn Release #1.0, Version #1.25
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/451,096
25     : FILING DATE:
26     : CLASSIFICATION: 435
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: US 08/219,842
29     : FILING DATE: 30-MAR-1994
30     : ATTORNEY/AGENT INFORMATION:
31     : NAME: Campbell, Cathryn A.
32     : REGISTRATION NUMBER: 31,815
33     : REFERENCE/DOCKET NUMBER: P-AG 9504
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: (619) 535-9001
36     : TELEFAX: (619) 535-8949
37     : INFORMATION FOR SEQ ID NO: 1:
38     : SEQUENCE CHARACTERISTICS:
39     : LENGTH: 1613 base pairs
40     : TYPE: nucleic acid
41     : STRANDEDNESS: double
42     : TOPOLOGY: linear
43
44 US-08-451-096-1

```

Query Match	23.7%;	Score 397.2;	DB 1;	Length 1613;
Best Local Similarity	58.1%;	Pred. No. 7.2e-99;		
Matches 743; Conservative	0;	Mismatches 523;	Indels 12;	Gaps 2;

QY	244	aacggcaacccctgtgaaacgcaatggttaaccctacacatggaatctcgaagaaagttctcttctgtg	303
QY	219	AACGCCAACATATACAAAGCTATCTGTACACACCCATCATTTGTAAATAAATCTTCTTCATA	278
Db	304	ggtatcccccgaaatgtctcgtgtgttttggtaacatctcgaagccgcgcgaataatccgcgcct	363
QY	279	GTAATAATCCCATATATATGGAAGCTTTTGCAACAGTACAGTATCCCTCAATAAATACGGTGCC	338
Db	364	ccgagatatgacctctccgcgcgatgaacaacctgcgtctcttgcgcgtgtcatctgcgcgttacc	423
QY	339	CCCGATTATGGCTTTTCCCCGATTAACAAACATAAGCTTCTACCTCTTAACTCCCTCTCTC	398

QY 424 gcatbaggcgtgcttcgctgctgcacacggcggtgacgtgctggttcggcgctt 483
 Db 399 CTACTCTGCTCTGCTATGAGGCGGAGACAGTATGAAACAGTCC--- 455
 QY 484 ggttggttcgtgacccgctgctgacccggaagctggtcattcgatgacctcgcg 543
 Db 456 -----TACCCTCCCTTAGAGGGAACCTACCTCCACCTGGACCCCTCCCTAGACCTAACCC 509
 QY 544 atttcgaggttcacgttgcgtgctgcctcctgacatggcgcgatcaacatgacga 603
 Db 510 ATCTTCCTTACCTAGAGAGTGTCTCTCTATCTTAGGSCCATCAATTTTCAATCACA 569
 QY 604 acccttctgaacacgacgaccccgacgtgacgtgacacaaagtgcgttgcgttg 663
 Db 570 ACATATATCAATATAAACCCTCCCTGGCCATACCCAAAGCCCTCTCGTCTGA 629
 QY 664 tgcattcttaccagcgttgcgtgacccgctggtggtggtggttgcgttgcgtgaac 723
 Db 630 TCGGTCTTATCAGACAGTCTTACTTCTCTATCTCTCCAGTCTGACCTCTGCGATC 689
 QY 724 accatgctgctgacacgacgttgcgtgacaccccttcacatcctgctggtgacgt 783
 Db 690 ACTATCTACTAAGACAGACCGCAACCTACACACCTCTTGACACCCCGGAGAGAGA 749
 QY 784 gaccgattcgttaccacacacatcctggttcttgggacccggaagtgtacatc 843
 Db 750 GACCCCATTTTATACCAACACCTATTTCTGATTTTTCGGTCCAGTCTGAATTTATTTCT 809
 QY 844 atctgacggcgttgcgtgacacacacacacacacacacacacacacacacacacac 900
 Db 810 ATCTCTACGAGGCTCGGATATATCTCCATTTTCTTACTTACTTCCGGAATAAAGAA 869
 QY 901 gcttcggttaccctgacgtgctgacgtgacgtgacgtgacgtgacgtgacgtgac 960
 Db 870 CCATTTGATATAGTATGATGCTGAGCTATGATATGATATGATATGATATGATATG 929
 QY 961 gcttgagcgac 1020
 Db 930 GTGTGACACACATATATTTACATGAGTAATGACATGACATGACATGACATGACAT 989
 QY 1021 ctgac 1080
 Db 990 TCCGCTACATATATGATGCTTCCACACCGGCTCAAAAGTTTATGCTGACTCCACCA 1049
 QY 1081 atgttgagcgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 1140
 Db 1050 CTCACGGAACAAATATGATGCTGCTGACATGCTGACATGCTGACATGCTGACATG 1109
 QY 1141 ctgttcac 1200
 Db 1110 CTTTTACCGTAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1169
 QY 1201 taccac 1260
 Db 1170 CTACACGACAGCTACTAGCTTATACCCACCTTCCACTATGCTCTATCAATAGGAGCTGTA 1229
 QY 1261 ttgacgac 1320
 Db 1230 TTTGGCATCATAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1289
 QY 1321 gaatgagcgtgac 1380
 Db 1290 CAACCTTACGACAAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1349
 QY 1381 ccgac 1440
 Db 1350 CCACACACATTTCTGCGCTATTCGAGATGCGCCGACCTTACTGAGGATACCCGATGCA 1409
 QY 1441 ttgacgac 1500
 Db 1410 TACACACATGAAACATCTTATCATCTGTAGGCTCATCTTCTTCTTCAACGACGATATA 1469
 QY 1501 ttcttcac 1518

Db 1470 TTAATAATTTTCAATCAT 1487

RESULT 3
 US-08-810-599-1
 ; Sequence 1, Application US/08810599
 ; Patent No. 5976798
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: PARKER, W. Davis
 ; APPLICANT: HERNSTADT, Corinna
 ; APPLICANT: GHOSH, Soumitra S.
 ; APPLICANT: FAHY, Eoin
 ; TITLE OF INVENTION: Methods for Detecting Mitochondrial Mutations
 ; TITLE OF INVENTION: Diagnostic for Alzheimer's Disease and Methods for Determin
 ; NUMBER OF SEQUENCES: 82
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Kenyon & Kenyon
 ; STREET: 1025 Connecticut Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: US
 ;
 ; ZIP: 20036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.25" floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Wordperfect 6.1 for Windows
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/810,599
 ; FILING DATE: Concurrent Herewith
 ; CLASSIFICATION: 436
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/757,438
 ; FILING DATE: 27 No. 5976798 1996
 ; APPLICATION NUMBER: US 08/614,072
 ; FILING DATE: 12 Mar 1996
 ; APPLICATION NUMBER: US 08/536,036
 ; FILING DATE: 29 Sep 1995
 ; APPLICATION NUMBER: US 08/414,969
 ; FILING DATE: 31 Mar 1995
 ; APPLICATION NUMBER: US 08/413,740
 ; FILING DATE: 30 Mar 1995
 ; APPLICATION NUMBER: US 08/410,658
 ; FILING DATE: 24 MARCH 1995
 ; APPLICATION NUMBER: US 08/397,808
 ; FILING DATE: 3 Mar 1995
 ; APPLICATION NUMBER: US 08/219,842
 ; FILING DATE: 30 MARCH 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Toffenetti, Judith L.
 ; REGISTRATION NUMBER: 39,048
 ; REFERENCE/DOCKET NUMBER: 2105/17
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-429-1776
 ; TELEFAX: 202-429-0796
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1613 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; HYPOHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-810-599-1

Query Match 23.7%; Score 397.2; DB 2; Length 1613;
 Best Local Similarity 58.1%; Pred. No. 7.2e-99;
 Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;


```

Db 6171 CCGGATATGCGGCTTCCCGCATTAACATAGCTTCTGACTTCTACCTCCTCTC 6230
Qy 424 ggcgtggcggtggtcgcgtcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 483
Db 6231 CTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6287
Qy 484 ggttgggttctgtacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 543
Db 6288 -----TACCCCTCCCTTAGAGAGGAACTACTCCACCTCGAGACCTCGTAGACCTAACCC 6341
Qy 544 atttgcgggttactctgtcgggtgcctcctcgcatacgcgcgcgcgcgcgcgcgc 603
Db 6342 ATCTTCTCTTACCTACAGAGGCTGCTCCTCTCTCTCTCTCTCTCTCTCTCTCT 6401
Qy 604 acccttctgaacatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 663
Db 6402 ACATATATCATATTAACACCCCTGCTATTAACCAATACCAACGCGCTCTCTCTCT 6461
Qy 664 tgcattctatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 723
Db 6462 TCCGCTCTATATACAGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6521
Qy 724 accatgtctgtacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 783
Db 6522 ACTATATCTACTAACAACAGCAGCAACCTCAACACACCTCTCTGACCCCGAGAGGA 6581
Qy 784 gaccgcattctgtacacacacacacacacacacacacacacacacacacacacac 843
Db 6582 GACCCCATTTCTATACCAACACCTATTTCTGATTTTTCGACACCTCGAATTATTTCTT 6641
Qy 844 atttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 6642 ATCTTCTCTTACAGGCTTGGATTAATCTCCCATTTGTAATCTACTCTCGGAAAAAAGA 6701
Qy 901 gcttcggttactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Db 6702 CCATTTGGATTAATAGTATGTTGCTGAGTATGATATATCAATTTGGCTTCTTATC 6761
Qy 961 gcttgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
Db 6762 GTGTGAGCAGACCATATATTTACAGTAGAGAAATAGACAGACAGACATTTTCAAC 6821
Qy 1021 ctcgcgcacatgtgtatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
Db 6822 TCCGCTACCATATATATGCTATCCCAACGCGCTCAAAAGTATTAGCTGACTCCGACA 6881
Qy 1081 atgtgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
Db 6882 CTCCACGGAGCAATATGAAATGATCTGCTGAGTCTCTGAGCCCTTAGATTCATCTTT 6941
Qy 1141 ctgttcaacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
Db 6942 CTTTTCACCGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7001
Qy 1201 taacacgaacattatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
Db 7002 CTACACGACAGCTACTAGCTTGTAGCCACATTCACATATGCTCTATATAGGAGCTGA 7061
Qy 1261 ttgtgcatcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1320
Db 7062 TTTGGCATATAGAGAGGCTTCACTTCACTGATTTTCCCTATTTCTAGCGGTACACCTTAGAC 7121
Qy 1321 gaatgggtgtcaaaagctgtacattctgtacacacacacacacacacacacacac 1380
Db 7122 CAACCTTAGCGCAAAATCATTTTCACTATATTTCACTGCGCGTAAATCTAAGTTTCTTC 7181
Qy 1381 ccgcgcacacacacacacacacacacacacacacacacacacacacacacacacac 1440
Db 7182 CCACAGACATTTCTGCGCTATCCGAGATGCCGAGACGTTACTCGGACTACCCGAGATGA 7241
Qy 1441 ttgcgcgtgtggaacaaagctcgttctatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500

```

```

Db 7242 TACACCAATGAACATCCTATCATCTGTAGGCTCATTTCTTCTATACAGAGTAATA 7301
Qy 1501 ttcttcacatgtatcttc 1518
Db 7302 TTAATATATTTTCATGATT 7319

RESULT 5
US-09-097-889-1
; Sequence 1, Application US/09097889
; Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: Herinstdt, Corina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,889
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.417
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-097-889-1

Query Match 23.2%; Score 388.2; DB 4; Length 6744;
Best Local Similarity 57.1%; Pred. No. 3.7e-96;
Matches 730; Conservative 19; Mismatches 517; Indels 13; Gaps 3;

```

```

Db 2862 -----TACCGTCCTTRGCAAGGAAGTAAGTCCACGTCGTCGTAAGCACTAAC 2915
OY atttcgcggtcactgttcggtgccc- tgcatacggcgcgatcaacatgatac 602
Db 2916 ATCTTSTCCCTTACACGTAGAGGTTTCCTTATCTTATAGGGGCAATATTCATC 2975
OY gacctcttgaaacagcgcccgatgacgtgacaaagtgcggtcttcgtc 662
Db 2976 AACATTTATATATATAAACCCCGCATATACCAATACCAAGCCCGCTTTCGTG 3035
OY gtcacatcttaacagcggttgatccgtgctgcgcgtccggtctcgtgtgcaat 722
Db 3036 ATCCGCTCAATACAGAGAGTCTACTTCTCTATCTCCAGTCCAGTCCAGTCCAGT 3095
OY 663 gtcacatcttgatcagcgacgttaacttgagagacgtcttcaatccgtctgagcgcg 782
Db 3096 CACTATACACTAACAGACCGGTACACACGACCTTTTTCGACCCGCGGAGAGG 3155
OY 783 tgaccgactctgtaacacacacatcgtgtctcttgagcaccggaagtgtacatcat 842
Db 3156 AGACCCCATTTATACACACACTATTCGATTTTCGTCACCTGAAAGTTATATCT 3215
OY 843 catctgcgcggtcttgatcatacaagcagtgctgtgcgaactctcgt---aaaaagc 899
Db 3216 YATCCCTACAGGCTTCGGAATATCTCCCATTTGTACTTACTCCCGGAAAAAAGA 3275
OY 900 gctcttcggttacgcgcgagtgatgcaatggtgcaatcggtgtcttcgtcgtctgt 959
Db 3276 ACCATTTGATATCATAGTATGCTGTGACCTATGATATCAATTTGGCTTCTGAGGTTTAT 3335
OY 960 cgtcttgagcagacacatgttacacggttggtatgtcgtgcgacgaatcctactcat 1019
Db 3336 CGGTGAGACACACATATTTTACAGTAGAATAGACGATAGACACAGCATATTTTAC 3395
OY 1020 gctgcgcacacagtgatcgcggtgcgcgcgacttaagctctcgtgtgacgcaac 1079
Db 3396 CTCCGCTACCATATCATGCTATCCACGCGCTCAAGATTTTACGTCTGACGCCAC 3455
OY 1080 gatgtgagcgcgctcggtgtgagttcaacacgcgcgactgcggtctgtgttact 1139
Db 3456 ACCTCCAGGAGCAATATGAATGATCTGCTGAGTCTGAGCCCTGAGATTTATTT 3515
OY 1140 cctgttcacggtgtgtgtgacgcgtatcgtgtcgcgcgaagcggtctgtgacggtgc 1199
Db 3516 TCTTTTCACCGTATGCTGCTGCTGACGATTTGATTAAGCAAACTCATCAGACATCGT 3575
OY 1200 atatacagacacatatacgtgtgacacatcaatgtatgtcgtgtgacgac 1259
Db 3576 ACTACAGAGACGTAAGTCTGATGACCACTTCCACTATGTCCTATCAATAGAGACTGT 3635
OY 1260 ctttgacatctgcgcggtatctacttctacatgcgaaagtctcgcgcgcgcttccc 1319
Db 3636 ATTTGTCATCATATGAGGCTTATCTACGATTTCCCTATTTCTCAGGGTACACCTAGA 3695
OY 1320 ggaatgggtgcgaagctgcactctgcgaactcttcaatcggtgtgcgaagtgacttct 1379
Db 3696 CCAAACTACGCAAAATTCATTTCCCTATTCATATTCATCGGGTAAATTCATCTTCT 3755
OY 1380 ccgcgcgcgttgagagcgatgcgacgtgcgcgcgtttacatgcacatccgaacg 1439
Db 3756 CCACACACACTTTCTCGGCTATCCGAAATGCCCAAGCTTACTCGGACTATCCGATGC 3815
OY 1440 cttcgcgctgtgaaacaaagtctcgtcctatgtgtgcttccgtcgtcgtcgtcct 1499
Db 3816 ATACACACATGAATAATCTATCATCTGTAGGCTCATTCATTTCTTAACAGCACTAAT 3875
OY 1500 gttcctcatcgtgacttt 1518
Db 3876 ATTAATTAATTTTCATTAAT 3894

```

RESULT 6
US-08-413-740A-1

```

: Sequence 1, Application US/08413740A
: Patent No. 6171859
: GENERAL INFORMATION:
: APPLICANT: HERRNSTADT, CORINNA
: APPLICANT: PARKER, WILLIAM D.
: APPLICANT: DAVIS, ROBERT
: APPLICANT: MILLER, SCOTT W.
: TITLE OF INVENTION: Diagnosis, Therapy and Cellular and
: TITLE OF INVENTION: Animal Models for Diseases Associated with Mitochondrial
: NUMBER OF SEQUENCES: 206
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenyon & Kenyon
: STREET: 1025 Connecticut Avenue, N.W.
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20036-5405
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/413,740A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04063
: FILING DATE: 30-MAR-1995
: APPLICATION NUMBER: 08/413,740
: FILING DATE: 30-MAR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Bonham, David B.
: REGISTRATION NUMBER: 34297
: TELEPHONE: (202) 429-1776
: TELEFAX: (202) 429-0796
: INFORMATION FOR SEO ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1735 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-413-740A-1

Query Match      23.0%; Score 384.4; DB 4; Length 1735;
Best Local Similarity 58.1%; Pred. No. 2.3e-95;
Matches 742; Conservative 0; Mismatches 521; Indels 15; Gaps 3;

```



```

QY 604 accctcttgaaacatgacgagcccgagatgaagctgcacaaagtccgtgtctcgtg 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 570 ACAATATCAATATTAACACCCCTGACCAATACCAATACCAACGCCCTCTGCTCA 629
QY 664 tgcatttatacagcgtgagctgaccccgagcgtgcgcgttcgtgctgtgcaac 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 630 TCCGCTTAATACACAGACTCTACTTCTCTATCTCTCCAGTCTTACTCTGTGCAATC 689
QY 724 accatgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 690 ACATATCTACTATACACACGACGACGACGACGACGACGACGACGACGACGACGAC 749
QY 784 gaccgattctgtaacacacacacacacacacacacacacacacacacacacacac 843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 750 GACCCCAATCTATACCAACACCTATCTATCTATCTATCTATCTATCTATCTATCT 809
QY 844 attctgcccgtcttgacatcatcaagcattgctgacacacacacacacacacacac 900
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 810 ATCTTACACAGGCTTCGGAATATCTCCATATTTGTAACCTTACTACCTCGGAAAAAAGA 869
QY 901 gctctggttactcgtccgacgtgctatgcaatggtggaatcgtgttctgagcttgc 960
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 870 CCATTTGATACATAGATAGTGTGATGATGATGATGATGATGATGATGATGATGATG 929
QY 961 gctcgtgacgacacacacacacacacacacacacacacacacacacacacacac 1020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 930 GTGTGACACACACATATATTTTACAGTAGAGATAGACCTAGACACACACACATATTTTAC 989
QY 1021 ctgagcacaatggtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 990 TCGGCTACCAATATATCATCGCTATGCCACGCGGTCAAGATTTTGTAGTACTGCGCAC 1049
QY 1081 atgtggtgagcgtcgtggtgacacacacacacacacacacacacacacacacacac 1140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1050 CTCACACGAGCAATATGAAATATGATCTGACAGTGTCTGACGCTTACGACCTTACGATTC 1106
QY 1141 ctgttcaacgtgagtggtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 1200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1107 CTTTTCACCGTATAGTGTGCTGACGCTGATGTATGTAACAACTATCAGTACAGCATGCTGA 1166
QY 1201 tatacagacacacacacacacacacacacacacacacacacacacacacacacac 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1167 CTACACGACACGATCTAGTGTGAGCCACCTTCCATCTATCAATAGAGAGCTGA 1226
QY 1261 ttgagcattcgtcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 1320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1227 TTTGCCATCATAGAGGCTTCATCTACTGATTTCCCTATTTTCAGGCTTACACCTTAGAC 1286
QY 1321 gaacggcgtgacacacacacacacacacacacacacacacacacacacacacacac 1380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1287 CAACACGACGACCAATATCATCTATCATATTTTCAGGCGTAATTAATCTTCTTC 1346
QY 1381 ccgacgacacacacacacacacacacacacacacacacacacacacacacacacac 1440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1347 CCACACACATTTCTCGGCTATTCGGAATGCCCGACGTAATCTGACCTACCCGATGCA 1406
QY 1441 ttgagcgtgagacacacacacacacacacacacacacacacacacacacacacacac 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1407 TACACACACATGAATCTATCTATCTGTAGGCTCATTTCTCTAAGACAGCATATA 1466
QY 1501 tctctacacgtgacacacacacacacacacacacacacacacacacacacacacac 1560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1467 TTAATTAATTTTCAATGATT 1484

```

RESULT 8
 US-08-998-416-564/C
 : Sequence 564, Application US/08998416
 : Patent No. 6239264
 : GENERAL INFORMATION:
 : APPLICANT: Philippesen, Peter
 : APPLICANT: Pohlmann, Rainer

```

APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corline
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PE/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEO ID NO.: 564:
SEQUENCE CHARACTERISTICS:
LENGTH: 773 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1390UP
US-08-998-416-564

```

Query Match 8.8%; Score 147.2; DB 4; Length 773;
 Best Local Similarity 51.5%; Pred. No. 6.9e-31;
 Matches 369; Conservative 0; Mismatches 338; Indels 9; Gaps 1;

```

QY 250 cacctgtgaaacatgacgttactcaacatgacgtatctgatagtgtctctgtggtalc 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 710 CAATTTTAATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 651
QY 310 ccgacgtgtggtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 650 CCAATTAATTAATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 591
QY 370 atgagctcccgatgaaacacacacacacacacacacacacacacacacacacacacacac 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 590 ATATCAATTTGCTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
QY 430 ggcgtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 539 TTAATTTGTTTATTAACATCTACTATTAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 490 gttctgtacccgctgctgcgaccccgagcgtgctatctgacgtgacgtgacgtgacgtgac 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 TATCTCTCTTATCTTCTATTAATCAATCAATCTGCTGCTTCTGCTGCTTATTAATTTTC 420
QY 550 gcggttcaactgtcgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 TCTTTACATTTTACTACTATTTCTTCTATTAATTAATTAATTAATTAATTAATTAATTAAT 360

```


TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 1138:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 719 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1692UP
 US-08-998-416-1138

Query Match 6.6%; Score 111; DB 4; Length 719;
 Best Local Similarity 51.1%; Pred. No. 4,9e-21;
 Matches 261; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 1008 atcctactcattcgtcgtccacatgtagtcggtgcccagcgacatgaatcctc 1067
 DB 15 acctatttttactcagctactatatttcttatttctactagatttaagctattag 74
 QY 1068 gtggatcgccagagatggggggcggtcggttgatgtaaaatcgccgagctcggcctt 1127
 DB 75 ttgattacttaactattttatgggttcattatagattactaacacacattatattctatt 134
 QY 1128 tggcttatgtccgttcacgctgggtggtggtccggtatcggtcgcccaagcgag 1187
 DB 135 atcatt 194
 QY 1188 tctgacgcgtgacatcacacacacattacgttggtgacacacacacacacacacacac 1247
 DB 195 attagatgtagatccatccatgatttatttatttatttatttatttatttatttattt 254
 QY 1248 gctgggtgacatccttgacatccttgacatccttgacatccttgacatccttgacat 1307
 DB 255 tttagtctgttatttcttatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 314
 QY 1308 ccgagcctcccggaatgggctgcaagctgacatccttgacatccttgacatccttgacat 1367
 DB 315 tttaatt 374
 QY 1368 cgtacgcttcccgcaagcattcctggagctgaggtatgctgcgagcgcgttaccaga 1427
 DB 375 ttt 434
 QY 1428 ctatccgaagcctcgcgctggtgaaacaaagctcgtcctatggtgcttccgctt 1487
 DB 435 ttatctctgattcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 494
 QY 1488 cgcctcgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1518
 DB 495 ttatatt 525

RESULT 13

US-08-998-416-289
 Sequence 289, Application US/08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jurgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Reibischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYYP11
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PP/5-30306/A/GC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 289:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 856 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1241UP
 US-08-998-416-289

Query Match 6.6%; Score 111; DB 4; Length 856;
 Best Local Similarity 51.1%; Pred. No. 5.3e-21;
 Matches 261; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

QY 1008 atcctactcattcgtcgtccacatgtagtcggtgcccagcgacatgaatcctc 1067
 DB 15 acctatttttactcagctactatatttcttatttctactagatttaagctattag 74
 QY 1068 gtggatcgccagagatggggggcggtcggttgatgtaaaatcgccgagctcggcctt 1127
 DB 75 ttgattacttaactattttatgggttcattatagatttctaacacacattatattctatt 134
 QY 1128 tggcttatgtccgttcacgctgggtggtggtccggtatcggtcgcccaagcgag 1187
 DB 135 atcatt 194
 QY 1188 tctgacgcgtgacatcacacacacattacgttggtgacacacacacacacacacacac 1247
 DB 195 attagatgtagatccatccatgatttatttatttatttatttatttatttatttattt 254
 QY 1248 gctgggtgacatccttgacatccttgacatccttgacatccttgacatccttgacat 1307
 DB 255 tttagtctgttatttcttatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 314
 QY 1308 ccgagcctcccggaatgggctgcaagctgacatccttgacatccttgacatccttgacat 1367
 DB 315 tttaatt 374
 QY 1368 cgtacgcttcccgcaagcattcctggagctgaggtatgctgcgagcgcgttaccaga 1427
 DB 375 ttt 434
 QY 1428 ctatccgaagcctcgcgctggtgaaacaaagctcgtcctatggtgcttccgctt 1487
 DB 435 ttatctctgattcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 494
 QY 1488 cgcctcgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1518

Db 495 TATATCATTAATCTTATCTTATATATAT 525

RESULT 14

US-08-518-878B-19
Sequence 19, Application US/08518878B

Patent No. 5702902

GENERAL INFORMATION:

APPLICANT: Tartaglia, Louis A.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESS:

ADDRESSER: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/518,878B

FILING DATE: 23-AUG-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-036

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 277 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-518-878B-19

Query Match 6.5%; Score 108; DB 1; Length 277;

Best Local Similarity 66.1%; Pred. No. 2.2e-20; Mismatches 80; Indels 0; Gaps 0;

Matches 156; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Db 661 tggatcgttatacagcgttgatcgtgctgctgctgctgctgctgctgctgca 720

1 TGATCCGACTATTACAGCGCTGCTCTCTATTATACAGTGCAGCGCAGGC 60

Db 721 atcacatgctgctgacagcgttaactcgagcagcactctcaactcgtgctgca 780

61 ATTACTATACACTAACAAGACCGCACTTAACCACTTTCTTGGCGTGGAGA 120

Db 781 ggtgaccgattcgtgacacacacatcctggtgcttcttgagcagcagagtgatc 840

121 GGGGACCAATTCCTACACAGCTGCTTCTGATTTCTTGGGCGCAGAGTTTATTT 180

Db 841 atcatctgcccggctgctgacatcagcagatgctgctgctgctgctgcaaaa 896

181 CTATCTCCCGAGATTGGAAATTATTTACATGTAGTACTACTCCGGAAA 236

Db 181 CTATCTCCCGAGATTGGAAATTATTTACATGTAGTACTACTCCGGAAA 236

RESULT 15

US-08-294-522B-19

Sequence 19, Application US/08294522B

Patent No. 5741666

GENERAL INFORMATION:

APPLICANT: Tartaglia, Louis A.

TITLE OF INVENTION: Compositions and Methods for the TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSER: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,522B

FILING DATE: 23-AUG-1994

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 277 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-294-522B-19

Query Match 6.5%; Score 108; DB 1; Length 277;

Best Local Similarity 66.1%; Pred. No. 2.2e-20; Mismatches 80; Indels 0; Gaps 0;

Matches 156; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Db 661 tggatcgttatacagcgttgatcgtgctgctgctgctgctgctgctgctgca 720

1 TGATCCGACTATTACAGCGCTGCTCTCTATTATACAGTGCAGCGCAGGC 60

Db 721 atcacatgctgctgacagcgttaactcgagcagcactctcaactcgtgctgca 780

61 ATTACTATACACTAACAAGACCGCACTTAACCACTTTCTTGGCGTGGAGA 120

Db 781 ggtgaccgattcgtgacacacacatcctggtgcttcttgagcagcagagtgatc 840

121 GGGGACCAATTCCTACACAGCTGCTTCTGATTTCTTGGGCGCAGAGTTTATTT 180

Db 841 atcatctgcccggctgctgacatcagcagatgctgctgctgctgctgcaaaa 896

181 CTATCTCCCGAGATTGGAAATTATTTACATGTAGTACTACTCCGGAAA 236

Search completed: December 22, 2001, 09:36:14
Job time: 5185 sec

Thu Dec 27 08:21:17 2001

us-09-712-768-1.rni

Db	86	ACGATCAAAATCTCAACACGTTAATGTCACCGCCACCGACTGCTTGTGATGATATTTCTTTATG	145
Qy	305	gtatcccgcatgtctcggtgttctgttaactactgtatgccccgtcaatcsggcgc	364
Db	146	TTATGGCTTATCATAAATTTGGAGGATTTTGGCAACTGATTTGTACCAATTAATATTTGGTGCAC	205
Qy	365	cgaataatgacctccgcgctatgaacaacctctgcttgctgtctcaatccgtacacg	424
Db	206	CCGACATAGCATTTCCACGGATTAATAATTAAGTTGCTGACTTT-----TACCC	255
Qy	425	cgaatggcgttgctctgctgttcgcaacggcggtgacggtcgaactgtgttcggcgctg	484
Db	256	CCATCTTTTCTTCTTCTACTAGCGTCTCCAAACAGTAGAGGAGGCGCAGAACGATGTA	315
Qy	485	gttgsggtctgtacccgcgcgtgtgcacccg-cgaagctgtgctatcatgtgacctgcg	543
Db	316	ACTGTGTACCTCTCATTTAGCTGTAACTGACCCACGCTGGGGCTTCTGTAGATTGACC	375
Qy	544	atttcgcggttacctgtctcggtgtgctctctcgatcaatggtgcgatacatgatcaag	603
Db	376	ATTTTTCCCTCCACCTACAGAGGTGATGATCATTTCTTAAGGGCATTAATTTCAATTACC	435
Qy	604	acctcttgaaactlgcgcgccccgcgcgaagcgcgcgaagaagtgcgtgtcttcctg	663
Db	436	ACAATTAATTAATTAATAAACACCACGCGCTATCTCAATATCAAAACCCCTCTATTGTTTGA	495
Qy	664	tcgaatcttatacagcgcttgctgtatcaatcgtctcgagcgctgtcgtctgtgtgtcaatc	723
Db	496	TCGTGTAATTAATTAACAGAGCTCTCTCTCTTTCATCAACCGCTGTAGCAGCTGTATT	555
Qy	724	accatgtctgtgcgcgacgtaactctgcgcgaactcttcaactcctgcgcgcgcgcgcgt	783
Db	556	ACAAATGCTTTAATCAAGATCAAAATTAACAAACAAATTTTGATCTCTCCGGAGGTGA	615
Qy	784	gaaccgcatctgtacaaacacacacccgtgtgtcttcttggtgcacccggaaagttaataatc	843
Db	616	GACCAATTTTATACCAACATCTTTCTGATTTCTTGGCACCCGAGATTTTCAATTTA	675
Qy	844	attctgcgcggtcttgatcatcatcaagccatgtctgtgcgaactcttgcg---aaaaagcgcg	900
Db	676	ATTCTCCAGGGTTTGGCATTAATCTCCCAATTTGTAACTTATTACTCAGTAAATAAAGAA	735
Qy	901	gtctcgttataccctgcgcgaatggtctatgtcaatgtgtgcgaatcgtgtcttggtcgttgc	960
Db	736	CCTTTTGGCTAATATAGCAATGCTGTAGCTAATAATATCATTTAGGGTCTCTAGGCTTATC	795
Qy	961	gtctggcgcaacaacatgtaacacgcttgatgtatgtcgtatgcccagcaatcctaactcatg	1020
Db	796	GTTTATGACACACCCACATATTTACAGTTTGGACTAGACGTTGATACCGGAGCTTACTACT	855
Qy	1021	ctgtagcaacatgtgtgatcgcgtgtgcgcgcgcgaatcaatcaactctctgtgtgactgcacg	1080
Db	856	TCCTGCCATGTATGATTATTTGCAATTTCCACAGAGATTAAGTATTAAGTTGATTAGCAACA	915
Qy	1081	atgtggggcggtctcgctgttgatgttcaaatcgcgcgaatgctcttgagcctlttgcttatgttc	1140
Db	916	CTCCATGAGGAACATCAATATGATATCCACAGCTTTTACTATGAGCCCTTGGATTATTTTTC	975
Qy	1141	ctgttataccgttgggtgtgtgtgacccgtatcgtgtgtgcccagaagcggtgtcgtgacccgtga	1200
Db	976	CTATTTTACAATCGGTGATTAACAGGATTTGTCTTAGCCCACTCTCATTAAGACATGCTTT	1035
Qy	1201	tatacgaacacattatcggtgtgtggcgcaactcatatgatagtatgtcgtgtgtgcgtc	1260
Db	1036	CTCCACGATACTTATTTATGTAGTAGCAACATTTCCATATGCTTATTCATATAGAGCTGTA	1095
Qy	1261	tttgagatcttcgcgcgtlatatacttllacatgacgaagttcctgggcgcgcgtctcccg	1320
Db	1096	TTTCGAATTAATAGGGGCTTCGTCCACTG-ATCCCCCTTTTACAGGTTATTCATTAATAT	1154
Qy	1321	gaatggagctgaagaagcgcaactcttgaaactcttcatcgtgtgcgaagctgaagctgtcttc	1380
Db	1155	GACTTAATGAGGAATTCACATTTCTCTATTTATTTTGTAGGTGTCAATTAATACCTTTCTTC	1214

[illegible]

RESULT	2
AU167717	
LOCUS	AU167717 1124 bp mRNA
DEFINITION	AU167717 Ol-br-ad cDNA Oryzias latipes cdna clone br0955, mRNA sequence.
ACCESSION	AU167717
VERSION	AU167717.1 GI:12589786
KEYWORDS	Esr.
SOURCE	Japanese medaka.
ORGANISM	Oryzias latipes

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Archivista: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:
Acanthomorpha: Acanthopterygii: Percomorpha: Atherinomorpha:
Belontiiformes: Adrianichthyidae: Oryziinae: Oryzias.
1 (bases 1 to 1124)
Mita,K., Ishikawa,I. and Yamauchi,M.
Establishment of cDNA database of medaka, *Oryzias latipes*
Unpublished (2001)
Contact: Mita K

Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method: uni-directional sequence direction: sequenced from 13 primer
(5' -> 3').

FEATURES	SOURCE	location/Qualifiers
		1..1124
		/organism="HNT"
		/strain="HNT"
		/db_xref="taxon:8090"
		/clone="br0951"
		/clone_1id="O1-br-ad cDNA"
		/sex="female/male mixed"
		/tissue_type="brain"
		/dev_stage="adult"
BASE COUNT		292 a 284 c 210 g 338 t
ORIGIN		

Query Match	18.4%	Score	307.8	DB	10	Length	1124
Best Local Similarity	59.8%	Pred. No.	5.4e-66				
Matches	533	Conservative	0	Mismatches	357	Indels	3
						Gaps	1

Accession	Sequence	Position
QY	683 ggcgtgacccgtctgagcgctgcacgggtctctgctcgtgltgcacatcaaccatgctctgacgacc	74
Db	66 TATTACTCTTCTTCTCTCTCTGCTGAGCTGCGAGGAGATCATATATCTTAACTGACAGCC	12
QY	743 gtaacttcggcgcgacacctctctcaatctctgtctcgtgcgcgcggcgtgagaccgattctgtgtaacaac	80
Db	126 GAACCTTAATATCAACCTTTCTTGCGACCTCGCAGGAGGAGGAGACCTTATTCATCCACAC	18
QY	803 aaatcctcgtgctctcttgtagcaccggagagtgatcatcatcatctctgcgcgatttggca	86
Db	186 ACCCTTTCTGATCTTTGGCCACCTGAGATTATATCTTAATCTTCCCGGGCTTGGAA	24

Query Match	17.0%;	Score 284.2;	DB 10;	Length 1235;
Best Local Similarity	61.0%;	Pred. No. 4.1e-60;		
Matches 479;	Conservative	0;	Mismatches 303;	Indels 3;
				Gaps

REFERENCE
AUTHORS
Phasiainmae; Galus.
1 (bases 1 to 821)
Abdrakhmanov, I., Lodygin, D., Gerotsh, P., Arakawa, H., Law, A., Plachy,
J., Korn, B., and Buerstedde, J. M.

Query Match	16.7 %	Score 279.6	DB 10	Length 866
Best Local Similarity	60.0 %	Pred. No 5,1e-9		
Matches 502	Conservative	0	Mismatches 330	Indels 5
				Gaps
Qy	623	ccccgcgcataagcctgcacaaagatgcctgtgtgtcttcctgtygtgcattcattcacagcctt	682	
Db	9	ccccccgacgtctacattaccacacacccccctatttcgatatccgcgtctcttcattatcccca	68	

REFERENCE	1 (bases 1 to 804)
AUTHORS	Abdrakhmanov, I., Lodygin, D., Gerolth, P., Arakawa, H., Law, A., Plal
TITLE	, J., Korn, B. and Buerstedde, J.M. A large database of chicken bursa ESTs as a resource for the
JOURNAL	analysis of vertebrate gene function
MEDLINE	Genome Res. 10 (12), 2062-2069 (2000)
COMMENT	20568495 Contact: Buerstedde JM

TITLE A large scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo
JOURNAL Development 128 (2001) In press
COMMENT Contact: Etensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: etensohn@andrew.cmu.edu.

FEATURES

source

1. 784

/organism="Strongylocentrotus purpuratus"

/db.xref="taxon:7668"

/clone="PC_0023_AL_H09_MR"

/clone.lib="Sea urchin primary mesenchyme cell CDNA library"

/tissue_type="embryo"

/cell_type="primary mesenchyme cells"

/lab_host="E.coli"

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo at priming from poly A+ RNA, directionally cloned"

BASE COUNT 184 a 184 c 157 g 256 t 3 others

ORIGIN

Query Match 16.3%; Score 273.4; DB 11; Length 784;
Best Local Similarity 60.8%; Pred. No. 1.7e-57;
Matches 461; Conservative 0; Mismatches 294; Indels 3; Gaps 1;

QY 718 gcaatacaccatgctgctgacacgacgtaacttcgcagacactcttcaactcgtgac 777
DB 23 GCCTCGGCATCTCTTCACAGATCGTAAATMAACACAACTTCTTGACCTGACGAGA 82
QY 778 ggcggtgacccgattctgtacacacalcctgcttctgggacccggaagtgac 837
DB 83 GGGGGAGATCAATCTGTTTCAACACTATCTGTTTGGACACCCCGAGGTGAT 142
QY 838 atataatctgcgcgacttgatcaltcaacgacatgctgctgcacactctgaaaaa 897
DB 143 ATCTTATCTTACCGGATTTGATGATCTCACAGTATACCTACACTCTGTGTAG 202
QY 898 c---cggtctcggttaactcgcagtgctatgcaatggtgcaatcggtctcgagc 954
DB 203 CGAGAGCCTTTCGATACCTGGGGATGTTATGCAATGCAATGAGATTTAGGA 262
QY 955 ttgctgctggcgacacacatgacacgctgtgcatgctgcgtgacccaacatctac 1014
DB 263 TTCCTGTCTGGGCCCATATGTTTACAGTAGAATGATGATGATACACAGCATAC 322
QY 1015 ttcattgctggcaccatgctgacgctgacgacgacgacgacgacgacgacgac 1074
DB 323 TTTCACGCGCCGACGATGATTTGCTTCCAAACAGGATTAAGGTTTACAGATG 382
QY 1075 gccaagatggtggcgctgctggtgacgacgacgacgacgacgacgacgacgac 1134
DB 383 GCAACGCTCCAAAGGCTTAATCTACATAGAGAACTCCCTATTATTAAGCTTGA 442
QY 1135 atcttctgctgacgctggtgctgacgctgacgctgctgacgacgacgacgacgac 1194
DB 443 GTATTTTATTTACATTAGAGACATCACAGGATTTGTTTGCATTTCTTCATTTG 502
QY 1195 cgtgacatacagacacactatcgtgctgacgacgacgacgacgacgacgacgac 1254
DB 503 GTTGTCTTATGATTAATCTACAGTGTGAGTACACTTTCATACGTTTTCATAGG 562
QY 1255 ggcacgttgcacgtcgcgcgtatctacttctacacgcaagctctcggcgcgct 1314
DB 563 GCTGTATTTGCAATCTTGGCTGTTTCACTGATGTTTCCCTCTTCTGTTATAGC 622
QY 1315 ttcccggaatggtgctgacgacgacgacgacgacgacgacgacgacgacgacgac 1374
DB 623 CTACACCCATTATGAGGAGAGGTTCACTTCTTCAATATGTTTGTGAGTCACTTAC 682

QY 1375 ttctcccgacacactctctggaagcaggtatgacgacgacgacgacgacgacgac 1434
DB 683 TTTTTCCTCAACACTTCTTAGTACGCGGNAATGCGACAGCTACGACTATCA 742
QY 1435 gaagctcgcgctggtggaacaaagctcgtctcctatg 1472
DB 743 GACGCTATACACTTGTGAATAATCTATCTNCTCAATTGG 780

RESULT 14

B1147837

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 823

/organism="Mus musculus"

/strain="FVB/N"

/db.xref="taxon:10090"

/clone="IMAGE:5053996"

/clone.lib="NCI_CGAP_L19"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 226 a 207 c 137 g 253 t

ORIGIN

Query Match 16.2%; Score 271.2; DB 11; Length 823;
Best Local Similarity 60.4%; Pred. No. 6.1e-57;
Matches 483; Conservative 0; Mismatches 313; Indels 4; Gaps 2;

QY 671 ttatacagcgttgctgactctgctgctgctgctgctgctgctgctgctgctgctgct 730
DB 14 TTATTAACGCCCTACTGCTCTTATATACACAGTGTGCTGACGCGACGATTAATAC 73
QY 731 tgcgtac 790
DB 74 TACTTAACACAGCAACCTTAACACACACCTTCTTGTGATCCGCTGAGAGGAGACCA 133
QY 791 tctgtac 850
DB 134 TTCTTACACAGATCTGTTCTTATTTCTTGGCGACCGAAGATTATTTCTTATCTCC 193
QY 851 ccgagcttgacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 907
DB 194 CAGGATTTGGAATTAATTCACATGATGATTAATTAATTAATTAATTAATTAATTA 253
QY 908 gttacctgcagatgctctatgcaatggtggaatcggtgctgctgctgctgctgctg 967

Thu Dec 27 08:21:24 2001

us-09-712-768-1.rst

Page 12